

# 비대면 국내 여행 산업 활성화 요인 분석

코로나 19로 인해 변화된  
국내 여행 및 소비 패턴 분석

2020년 문화관광 빅데이터 분석대회

DATA89



# INDEX

1. 주제 선정
2. 데이터 Data
3. 데이터 분석
4. 데이터 분석 결과
5. 결론 및 시사점

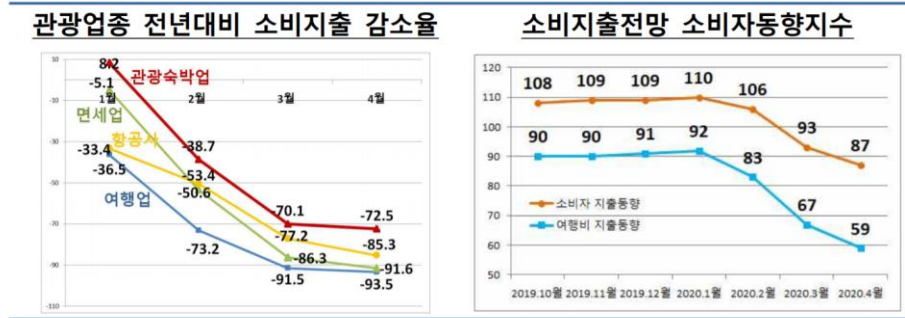


# 1. 주제 선정 – 문제 제기 및 연구 배경

## 여행비 지출은 감소했지만 소비자들의 국내 여행에 대한 높은 수요

### • 코로나19로 인해 줄어든 관광 수입

1)

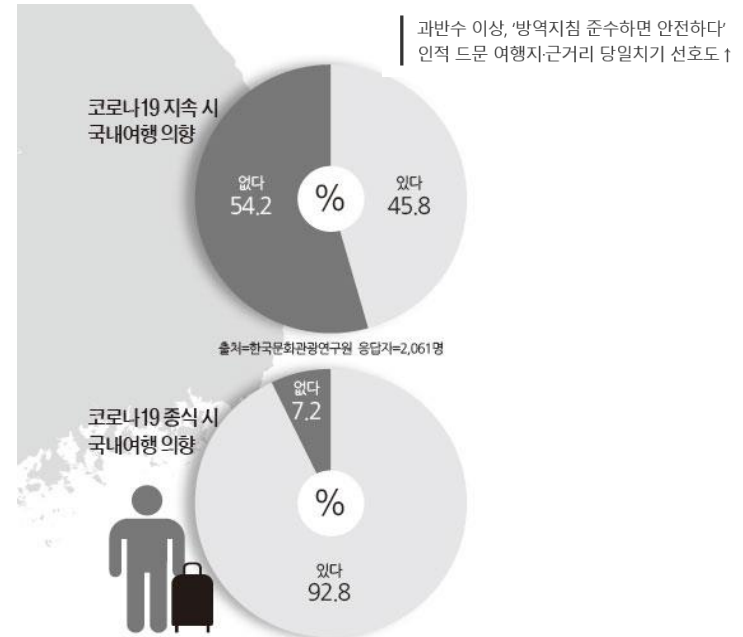


(서울=연합뉴스) 김보경 기자 = 신종 코로나바이러스 감염증(코로나19) 확산으로 관광업계가 최악의 위기에 빠진 가운데 올해 3월 우리나라의 관광 수입이 9년래 최저 수준으로 쪼그라든 것으로 나타났다.

12일 한국문화관광연구원 관광지식정보시스템에 따르면 코로나19가 전국에 창궐한 올해 3월 우리나라 관광 수입과 관광 지출은 각각 7억2천980만 달러(8천904억 원), 8억8천150만 달러(1조754억 원)로 잠정 집계됐다.

특히 관광 수입은 2011년 1월 이후 9년여 만에 가장 작은 수준으로 급감했다.

### • 그러나 국내 여행에 대한 높은 수요



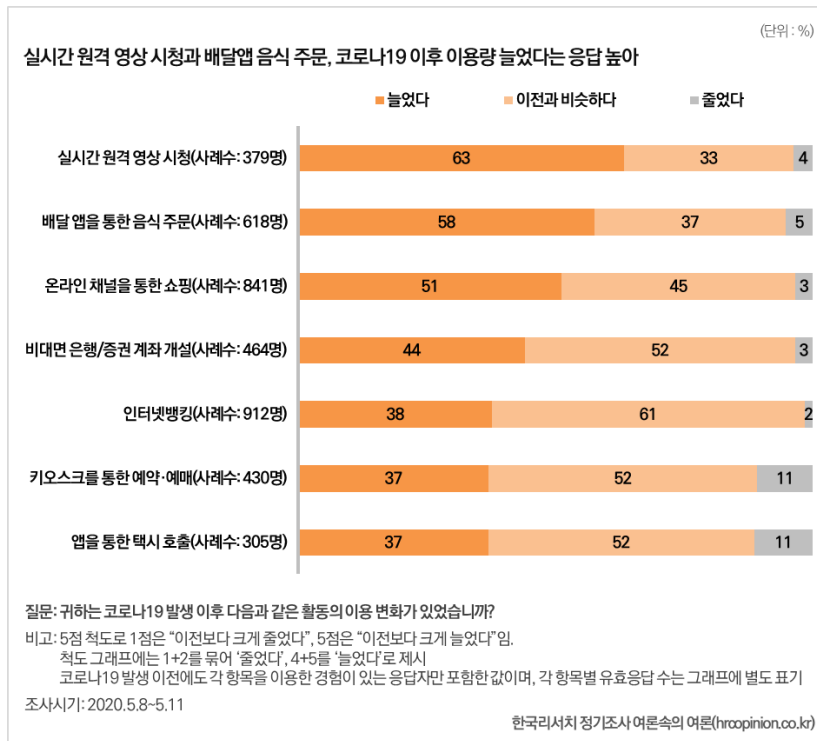
문화관광연구원 국내여행 조사결과 보니 -  
응답자의 46%  
“코로나 지속돼도 국내여행 가겠다”

# 1. 주제 선정 – 문제 제기 및 연구 배경

## 코로나 확산으로 “비대면”으로 변하는 사회 트렌드

- 코로나19로 인해 비대면 서비스 수요 증가

### 코로나19 이후 비대면 서비스 이용 빈도 늘어나



- 변화된 관광 생활 키워드



◆ 코로나19 여행 키워드 '안전'= 관광공사는 빅데이터 분석 결과를 토대로 코로나19 상황 속 여행 트렌드를 'SAFETY(안전)'로 정리했다. ▲근거리(Short distance) ▲야외활동(Activity) ▲가족단위(Family) ▲자연친화(Eco-area) ▲인기 관광지(Tourist site) ▲관광 수요회복 조짐(Yet...아직) 등 6개 키워드의 영문 앞글자를 묶은 것이다.



## 1. 주제 선정 – 해결 목표

해결 목표



코로나19로 부터 안전한 국내여행 제안을 통한 관광 활성화

## 2. 데이터 분석 – 분석 방향



- 국민여행조사
- 신한카드 데이터

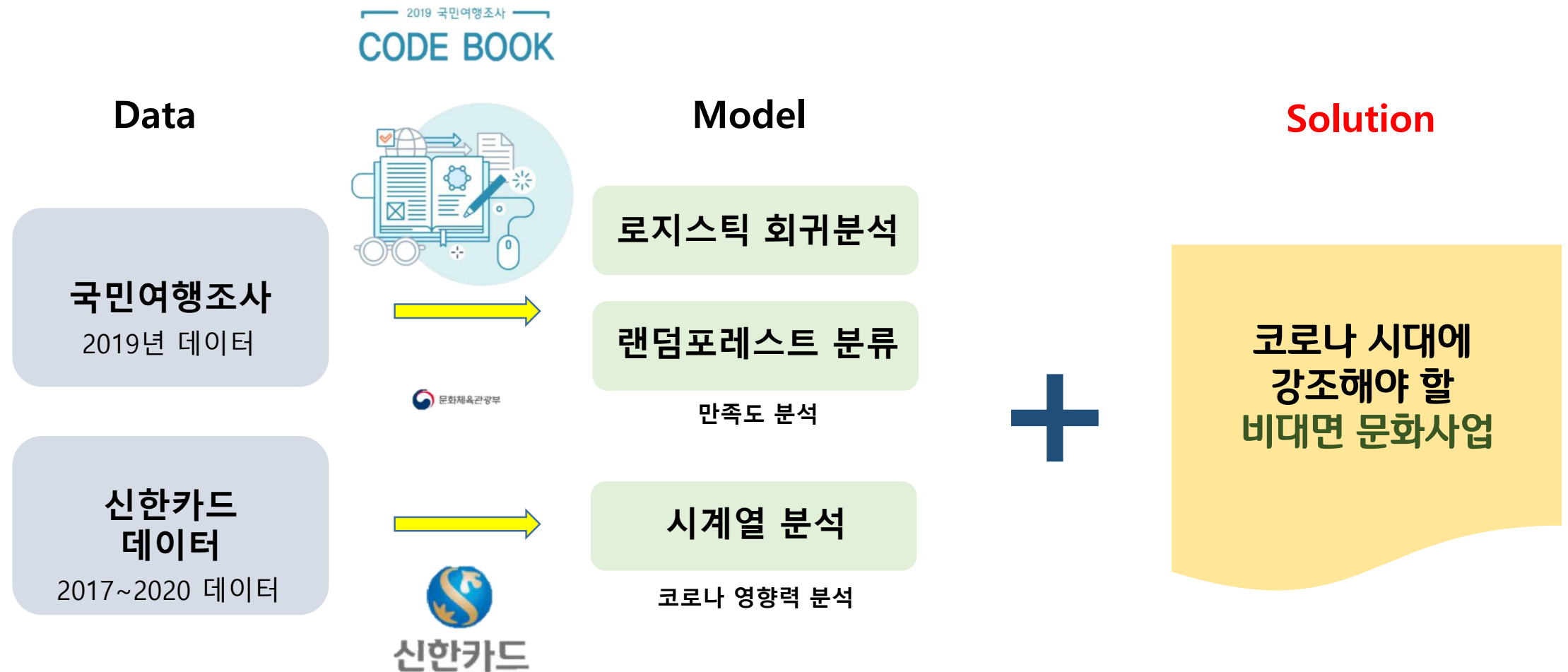
- 결측치 처리
- 데이터 스케일링
- 데이터 범주화

- 로지스틱 회귀분석
- 랜덤포레스트 앙상블
- 시계열분석

- 모델의 적합성 검증

- 코로나 시대에 맞는 여가 변수 추출

## 2. 데이터 Data – 분석 데이터 선정





## 2. 데이터 Data – 모델 설명

### Logistic Regression

- 여행지를 만족도로 분류하기 위해 선택한 모델
- $p(y|x) = \text{Bin}(y; \mu(x), N)$
- 종속변수가 이항 분포를 따르도록 만족도 변수를 조절하고, 이러한 변수가 여행 데이터의 비용, 위치 등 독립 변수들에 의해 의존한다고 가정하고 분석을 진행한다.
- 로지스틱 회귀분석에서는 시그모이드 함수를 사용하여 종속변수의 예측치를 가지고 값을 분류하게 된다.

### Decision tree Classification

- 여행의 만족도의 군집을 보여주기 위해 선택한 모델, 앙상블 모델인 Random Forest를 활용할 경우의 정확도가 더 높을 수 있지만, 하나의 모델을 이용하여 만족 고객 군집을 뽑아 내기 위해 이용하였다.
- 의사결정나무는 여러 가지 규칙을 순차적으로 적용하면서 독립 변수 공간을 분할하는 경우 사용하는 분류 모형으로 여기에서는 만족도를 종속변수로 갖기 때문에 분류 모델로 사용하였다.

### Random Forest Classification

- 랜덤 포레스트는 의사결정 나무의 앙상블 방법론으로 데이터 특정 차원의 일부만 선택해서 사용하며, 독립 변수의 차원을 무작위하게 감소시킨 다음 변수를 선택하게 된다. 이를 통해 과적합을 방지할 수 있고, 개별 모형의 성능이 낮을 경우 결합 모형의 경우 성능이 더 향상될 수 있다.
- 랜덤 포레스트는 취합 aggregation 방법론으로 부스팅 모델도 확인하여 보았지만 결과가 나은 취합 모델을 사용하기로 한다.

### Time series analysis

- 신한카드 데이터의 경우 변수들이 대부분 범주형이고, 이에 따른 매출액과의 명확한 상관관계를 파악할 수 없었다.
- 이를 활용하여 코로나 이후로의 데이터를 분석하기 위해 매출액이라는 순서열을 앙상블 평균을 통하여 구하는 방식을 이용하였다.
- Autoregressive 모형을 통해 정상성을 확인하려고 노력하였으며,
- ARIMA를 이용하여 시계열 분석을 진행하였다.



## 2. 데이터 Data – 데이터 구성

### 국민여행조사

2019년 데이터

기존 변수	변수 설명
Q	Q_타 지역(국내 및 해외) 방문 횟수
D_TRA1_YEAR	B_여행1차_시작날짜_년
D_TRA1_SMONT H	B_여행1차_시작날짜_월
D_TRA1_SDAY	B_여행1차_시작날짜_일
D_TRA1_EYEAR	B_여행1차_끝날짜_년
. . . .	
D_TRA1_COST	B_여행1차_여행 총경비
D_TRA1_NUM	B_여행1차_지출비용에 포함되는 인원수
D_TRA1_ONE_COST	B_여행1차_1인지출비용
D_TRA1_VAC	B_여행1차_평일이 포함된 경우 휴가/방학 여부

. . . .

사용 변수	변수 설명	데이터 설명
A12	Q_12_전반적만족도	Binaray로 변환 (만족, 매우만족 = 1, 보통, 불만족, 매우 불만족 = 0 )
date	B_여행1차_시작날짜	나뉜 데이터 YYYYMMDD 로 통합
Season	B_여행1차_계절	계절 별로 만족도를 확인해가 위함 ( 3월 4월 5월 = 봄, 6월 7월 8월 = 여름 , 9월 10월 11월 = 가을, 12월 1월 2월 = 겨울 )
D_TRA1_S_Day	여행1차_여행일수(박)	마지막 날짜를 계산할 수 있으므로 처리
D_TRA1_NUM	여행1차_지출비용에 포함되는 인원수	가족 단위 여행 파악 가능
D_TRA1_COST	B_여행1차_여행 총경비	지역경제 활성화의 중요 변수
D_TRA1_ONE_COST	B_여행1차_1인지출비용	지역경제 활성화의 중요 변수
D_TRA1_1_SPOT1	B_여행1차_1번째방문지	결측치가 적은 데이터를 선택

## 2. 데이터 Data – 데이터 전처리

### 국민여행조사 2019년 데이터

#### • 데이터 결측치 조정

```
1 # 결측치 확인
2 df_re.isnull().sum()

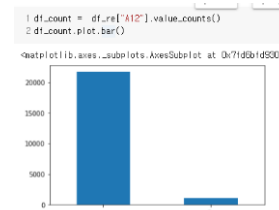
D_TRA1_CHECK      2068
D_TRA1_1_YEAR      0
D_TRA1_1_MONTH     0
D_TRA1_1_DAY       0
D_TRA1_1_YEAR      0
D_TRA1_1_MONTH     0
D_TRA1_1_DAY       0
D_TRA1_1_S_Day     0
D_TRA1_1_COST      0
D_TRA1_1_NUM       0
D_TRA1_1_ONE_COST   0
D_TRA1_1_VAC       15762
D_TRA1_1_TS2C      22687
D_TRA1_1_CASE      0
D_TRA1_1_1_YEAR    0
D_TRA1_1_1_MONTH   0
D_TRA1_1_1_DAY     0
D_TRA1_1_1_YEAR    0
D_TRA1_1_1_MONTH   0
D_TRA1_1_1_DAY     0
D_TRA1_1_1_SPOT1   0
A12                0
dtype: int64
```

- 설문지의 경우 많은 결측치 존재
- 통일성을 위하여 최대한 결측치가 없는 변수 위주 추출

#### • 명확한 분석을 위한 종속변수 범주화

```
1 df_re['A12'].replace(['만족', '매우 만족'], 1, inplace = True)
2 df_re['A12'].replace(['보통', '불만족', '매우 불만족'], 0, inplace = True)
```

#### • 분석을 위한 데이터 슬라이싱



\* 데이터의 만족/불만족의 데이터 불균형이 심각하여 분석을 위해 만족 데이터 랜덤 분할

train data size : (2281, 20)  
test data size : (978, 20)

```
df_train_1, df_test_1 = train_test_split(df_1, # 설명변수 데이터
                                         test_size = 0.9, # test 데이터의 비율
                                         random_state = 1234) # random state
```

#### • 다중 공선성 확인을 위한 scatter plot & VIF



	variable	VIF
5	D_TRA1_1_SPOT1	1.009000e+00
6	date	1.031000e+00
1	D_TRA1_S_Day	1.489000e+00
3	D_TRA1_NUM	1.771000e+00
4	D_TRA1_ONE_COST	2.851000e+00
2	D_TRA1_COST	3.051000e+00
0	const	1.198186e+09



### 3. 데이터 분석 – 로지스틱 회귀분석

#### 데이터 분석

회귀식 :

$A12 \sim \text{date} + C(\text{season}) + D\_TRA1\_S\_Day + D\_TRA1\_NUM + D\_TRA1\_ONE\_COST + C(\text{city})$

변수명	변수 설명	coef	유의 확률
city	B_여행1차_방문_시도	2.3337 ~ -1.7330	0.723 ~ 0.000
date	B_여행1차_시작날짜	-0.0471	0.000
Season	B_여행1차_계절	6.0433 ~ -28.2326	0.000
D_TRA1_S_Day	여행1차_여행일수(박)	0.1688	0.168
D_TRA1_NUM	여행1차_지출비용에 포함되는 인원수	0.1542	0.000
D_TRA1_ONE_COST	B_여행1차_1인지출비용	3.832e-06	0.000
Intercept	Intercept	9.507e+05	0.000

#### 유의미한 변수

##### 양의 상관 관계

겨울, 전북, 강원, 광주, 부산, 인원수, 1인당 지출 비용

##### 음의 상관 관계

봄, 여름, 대전, 날짜

### 3. 데이터 분석 – 랜덤 포레스트

#### 국민여행조사 2019년 데이터

#### Data

- 로지스틱 회귀분석과 동일하게 전처리



- 변수 범주화 후 대부분 데이터 이용

#### Default 가중치로 Classifier 모델 생성

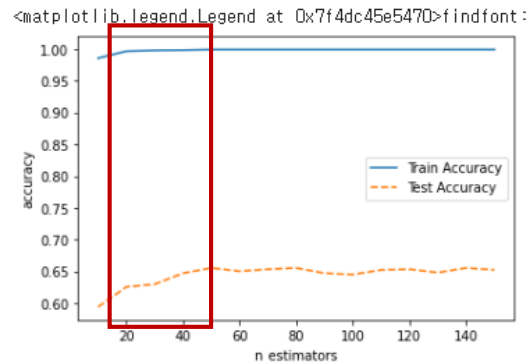
```
1 # 랜덤 포레스트 모델 생성: RandomForestClassifier
2 rf_uncustomized = RandomForestClassifier(random_state=1234)
3 rf_uncustomized.fit(df_train_x, df_train_y)
4 # train 데이터 셋 정확도
5 print("Accuracy on training set: {:.3f}".format(rf_uncustomized.score(df_train_x, df_train_y)))
6 # test 데이터 셋 정확도
7 print("Accuracy on test set: {:.3f}".format(rf_uncustomized.score(df_test_x, df_test_y)))
```

Accuracy on training set: 0.999  
Accuracy on test set: 0.645

과적합된 데이터 확인, customizing을 통한  
모델 정확도 향상 작업

파라미터 조정을 통한  
최적 RF Model 탐색

n\_estimator : 80



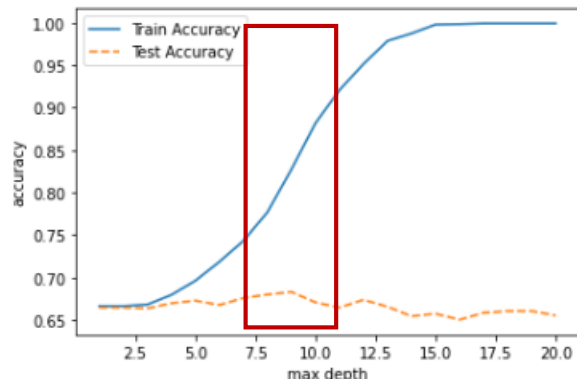
	n_estimators	TrainAccuracy	TestAccuracy
0	10	0.986	0.594
1	20	0.996	0.626
2	30	0.998	0.630
3	40	0.998	0.647
4	50	0.999	0.655
5	60	0.999	0.650
6	70	0.999	0.653
7	80	0.999	0.655
8	90	0.999	0.647
9	100	0.999	0.645
10	110	0.999	0.652
11	120	0.999	0.653
12	130	0.999	0.648
13	140	0.999	0.655
14	150	0.999	0.652



### 3. 데이터 분석 – RF 모델 성능 평가

max\_depth: 최대 깊이 변경

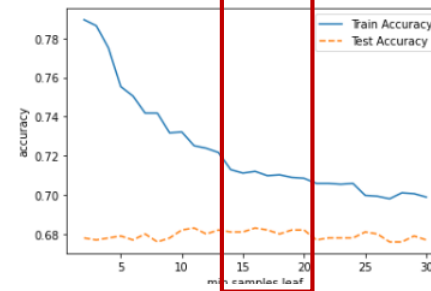
<matplotlib.legend.Legend at 0x7f4dc40e6c18>



	Depth	TrainAccuracy	TestAccuracy	
	0	1	0.666	0.665
	1	2	0.666	0.665
	2	3	0.668	0.664
	3	4	0.680	0.670
	4	5	0.696	0.673
	5	6	0.719	0.668
	6	7	0.743	0.676
	7	8	0.777	0.680
	8	9	0.827	0.683
	9	10	0.882	0.671
	10	11	0.921	0.665
	11	12	0.951	0.674
	12	13	0.979	0.666
	13	14	0.987	0.654

min\_samples\_leaf: 최소 잎사귀 수 제한

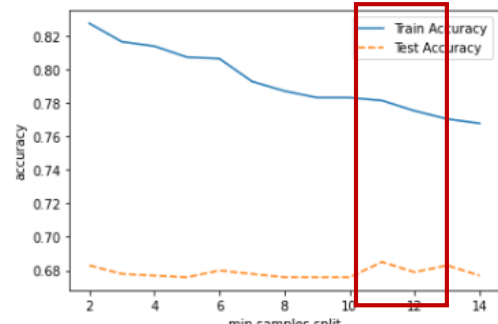
<matplotlib.legend.Legend at 0x7f4dc3f01a90>



	MinSamplesLeaf	TrainAccuracy	TestAccuracy
0	2	0.790	0.678
1	3	0.786	0.677
2	4	0.775	0.678
3	5	0.755	0.679
4	6	0.751	0.677
5	7	0.742	0.680
6	8	0.742	0.676
7	9	0.732	0.678
8	10	0.732	0.682
9	11	0.725	0.683
10	12	0.724	0.680
11	13	0.722	0.682
12	14	0.713	0.681

min\_samples\_split: 분할하기 위한 노드의 최소 샘플 수

<matplotlib.legend.Legend at 0x7f4dc3f894e0>



	MinSamplesSplit	TrainAccuracy	TestAccuracy
0	2	0.827	0.683
1	3	0.816	0.678
2	4	0.814	0.677
3	5	0.807	0.676
4	6	0.806	0.680
5	7	0.793	0.676
6	8	0.787	0.676
7	9	0.783	0.676
8	10	0.783	0.676
9	11	0.781	0.685
10	12	0.775	0.679
11	13	0.770	0.683
12	14	0.768	0.677

#### 최종 파라미터

min\_samples\_leaf = 10  
min\_samples\_split = 6  
max\_depth = 9  
n\_estimators = 80

### 3. 데이터 분석 – RF 모델 성능 평가

#### 최종 모델

```
# 최종 모델
2 rf_final = RandomForestClassifier(min_samples_leaf = 10, min_samples_split = 6, max_depth = 9, n_estimators = 80,
3                                   random_state=1234)
1 rf_final.fit(df_train_x, df_train_y)
5 # 평가
```

#### 평가

```
Accuracy on training set: 0.999
Accuracy on test set: 0.645

Accuracy on training set: 0.732
Accuracy on test set: 0.682
```

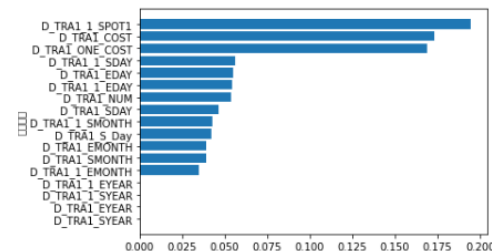
과적합을 줄이고, 기존 평가 모델의 성능 개선

#### 설명 변수 중요도 추출

	Feature	Importance
16	D_TRA1_1_SPOT1	0.194
7	D_TRA1_COST	0.173
9	D_TRA1_ONE_COST	0.169
12	D_TRA1_1_SDAY	0.056
5	D_TRA1_EDAY	0.055
15	D_TRA1_1_EDAY	0.054
8	D_TRA1_NUM	0.054
2	D_TRA1_SDAY	0.047
11	D_TRA1_1_SMONTH	0.043
6	D_TRA1_S_Day	0.042
4	D_TRA1_EMONTH	0.039
1	D_TRA1_SMONTH	0.039
14	D_TRA1_1_EMONTH	0.035
10	D_TRA1_1_SYEAR	0.000
3	D_TRA1_EYEAR	0.000
13	D_TRA1_1_EYEAR	0.000
0	D_TRA1_SYEAR	0.000

#### 만족도 중요도 변수

1순위 : 지역  
2순위 : 가격  
3순위 : 날짜





### 3. 데이터 분석 – (추가) 의사 결정 모델을 통해 만족도 군집 확인

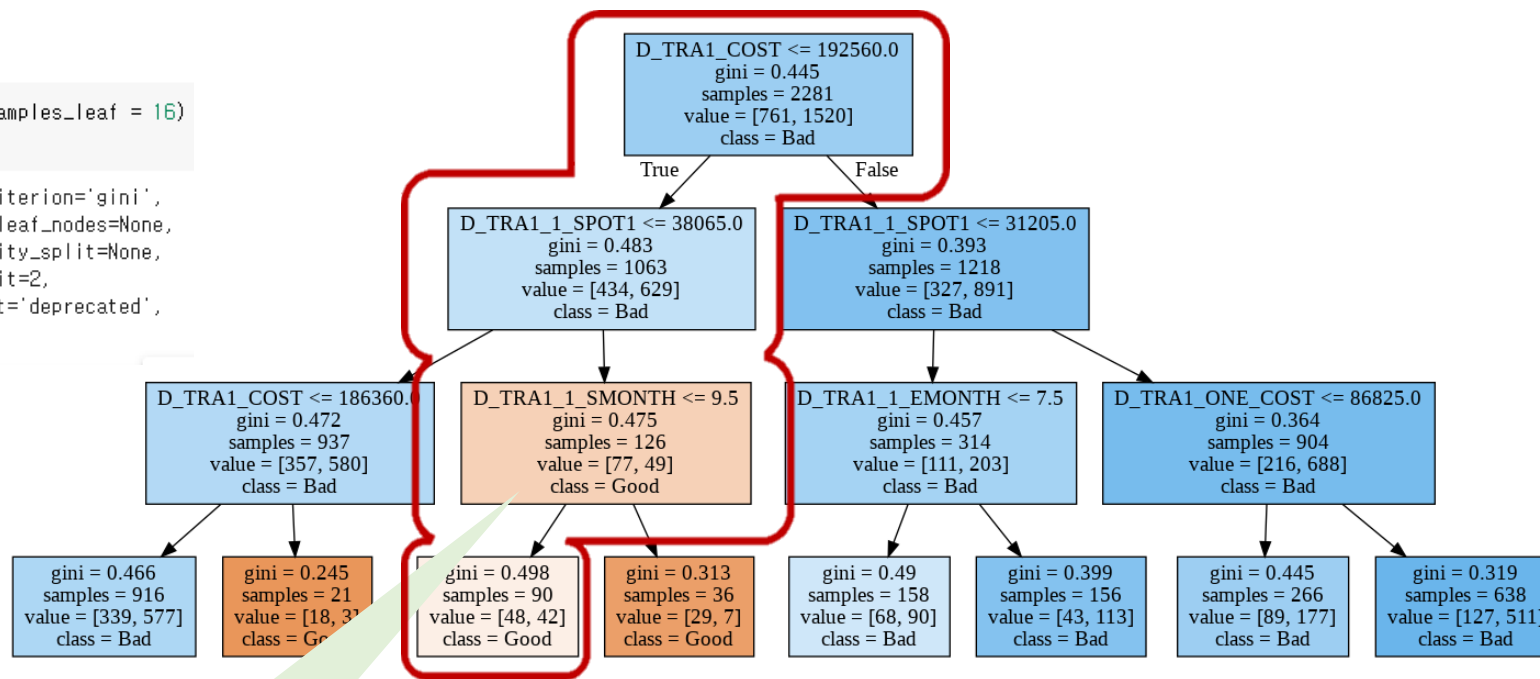
#### 의사결정트리를 이용하여 군집을 파악

RF와 동일하게 Customizing 진행

최종 모델 선정 / 시각화

```
1 tree_final = DecisionTreeClassifier(max_depth = 3, min_samples_leaf = 16)  
2 tree_final.fit(df_train_x, df_train_y)
```

```
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='gini',  
max_depth=3, max_features=None, max_leaf_nodes=None,  
min_impurity_decrease=0.0, min_impurity_split=None,  
min_samples_leaf=16, min_samples_split=2,  
min_weight_fraction_leaf=0.0, presort='deprecated',  
random_state=None, splitter='best')
```



가격이 192560원 이하이며, 위치는 38065를 초과하며, 시작일이 9월 인 경우 국내 여행에 만족한다.

### 3. 데이터 분석 – 만족도가 높은 상위 데이터 추출

#### 만족도가 높은 10개 시

```
1 df_importance.head(10)
```

	D_TRA1_1_SPOT1
37020.0	702
21090.0	663
39020.0	630
32060.0	620
36020.0	603
32030.0	545
35010.0	415
34030.0	391
37010.0	384
31370.0	380

경상북도 경주  
부산광역시 해운대구  
제주특별자치도 서귀포시  
강원도 철원군  
전라남도 여수시  
강원도 강릉시  
전라북도 전주시  
충청남도 보령시  
경상북도 포항시  
경기도 평택시



경주 관광지 : 동궁과 월지

관광지로 유명한 위치들이 만족도가 높은 상위 10개 도시에 있다.  
위와 같은 도시들은 경주, 해운대 등 한국 관광객 뿐만 아니라 해외에서도 아름다운 전경으로 인해 인기가 있는 곳이다.  
이러한 만족도 지표를 여행 산업 활성화에 이용할 수 있을 것이다.

#### 만족도가 낮은 10개 시

	D_TRA1_1_SPOT1
제주특별자치도 서귀포시	39020.0 30
전라남도 순천시	36030.0 22
경상북도 경주시	37020.0 22
전라남도 여수시	36020.0 20
경상남도 김해시	38070.0 19
경상북도 포항시	37010.0 19
강원도 강릉시	32030.0 18
제주특별자치도 제주시	39010.0 17
전라남도 신안군	36480.0 16
경상남도 남해군	38350.0 16



순천 관광지 : 순천만

만족도를 개수기준으로 분류하여  
만족도가 높은 도시가 만족도가 낮은 10개의  
시에 포함되기도 하였다.  
만족도가 높은 10개 도시에 속하지 않으며, 만  
족도가 낮은 도시들은 만족도 개선을 통한 지  
역 이미지 개선이 필요하다.



### 3. 데이터 분석 – 만족도가 높은 상위 데이터 추출

만족도가 높은 비용 상위 10개 |

```
1 df_count.head(10)
```

	D_TRA1_COST
300000.0	1649
100000.0	1623
200000.0	1604
150000.0	1341
500000.0	720
400000.0	719
50000.0	689
250000.0	524
120000.0	366
600000.0	345

만족도가 낮은 비용 상위 10개

	D_TRA1_COST
100000.0	90
200000.0	58
300000.0	50
150000.0	47
50000.0	41
120000.0	26
400000.0	26
500000.0	24
250000.0	18
80000.0	18

금액을 범주화 하여 분석을 진행하여 만족도가 낮고, 높은 비용에 겹치는 부분이 많다.

하지만 겹치지 않는 변수를 통해 사람들의 금액에 따른 만족도를 추측해 볼 수 있다. 최근의 트렌드는 여행을 떠날 때 가성비를 따지기 보다는 돈을 쓰고 즐기기 위해서라는 목적이 크다.

이러한 사람들의 여행 패턴도 분석을 통해 확인해 볼 수 있다.

2019년에도 패키지보다는 자유여행, '가심비, 소확행' 트렌드는 계속

이주상 기자 jsfan@chosun.com

기사입력 2019.03.08

최근 소비 트렌드로 자리를 잡고 있는 가심비(가격 대비 심리적 확실한 행복 추구) 붐은 올해 여행 분야에도 영향을 주고 있다. '고사'에 따르면 우리나라 사람은 해외여행을 갈 때 틈에 짜여진 선호하는 것으로 나타났다. 이것은 가심비와 소확행을 중시하는 다.



티몬이 최근 3개월간 매출을 조사한 결과 20대 사이에서 가심비를 중시하는 일점호화 소비 트렌드가 보이고 있다고 분석했다

## 2. 데이터 Data – 데이터 구성

### 신한카드 데이터

2017년 1월부터  
2020년 6월 까지의 데이터

변수명	변수 설명
V1	이용회원 거주지(17개 시도)
V3	카드이용 가맹점 주소(17개 시도)
V3	카드이용 가맹점 주소(시군구)
GB3	가맹점 업종 대분류 (문화예술, 스포츠, 여행, 취미)
GB4	가맹점 업종 세분류
SEX_CCD	성별
CLN_AGE_R	연령대(20,30,40,60,80)
TA_YM	이용년월
APV_TS_DL_TM_R	이용시간대(활동, 휴식, 취침)
VLM	취급액(단위: 원)
USEC	이용건수(단위 : 건)

사용 변수	변수 설명	데이터 설명
VLM	취급액(단위: 원)	종속 변수로 이용된다.
GV3	가맹점 업종 대분류	국내 여행 활성화를 목적으로 하므로 대분류 중 여행 만 추출
GV2	가맹점 업종 세분류	여행의 세분류로는 교통, 숙박, 관광쇼핑, 여행사, 체험
TA_YM	이용년월	시계열 분석을 위한 년,월 데이터 (201701~ 202006의 데이터)
코로나 변화 시점	2020.02	코로나19의 유행 이후 데이터 비교를 위해 설정

## 2. 데이터 Data – 신한 카드 데이터 전처리

### 신한카드 데이터

2017년 1월부터  
2020년 6월 까지의 데이터

### 데이터 전처리

여행에 관련된 정보만 추출하여  
가장 코로나 이후 매출에 영향을 많이 받은 변수 추출  
[ 시계열 분석 수행 ]

```
df_raw[df_raw['gb3']=='여행'][['gb2']].unique()
array(['숙박', '교통', '관광쇼핑', '여행사', '체험'], dtype=object)
```

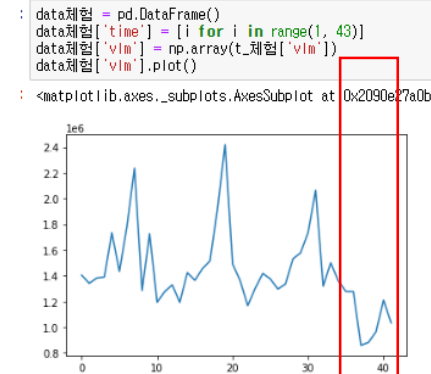
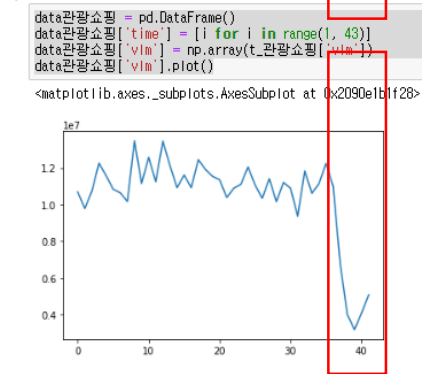
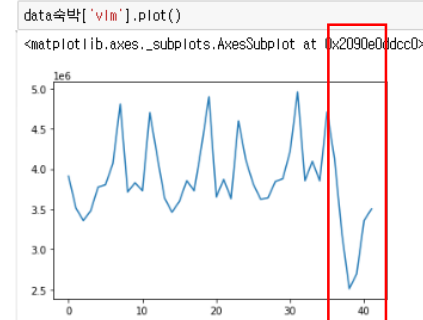
```
In [25]: df_숙박 = df_raw[df_raw['gb2']=='숙박']
grouped = df_숙박['v1m'].groupby(df_숙박['ta_ym'])
t_숙박 = pd.DataFrame(grouped.mean())
t_숙박
```

```
df_교통 = df_raw[df_raw['gb2']=='교통']
grouped = df_교통['v1m'].groupby(df_교통['ta_ym'])
t_교통 = pd.DataFrame(grouped.mean())
t_교통
```

```
In [25]: df_관광쇼핑 = df_raw[df_raw['gb2']=='관광쇼핑']
grouped = df_관광쇼핑['v1m'].groupby(df_관광쇼핑['ta_ym'])
t_관광쇼핑 = pd.DataFrame(grouped.mean())
t_관광쇼핑
```

```
In [26]: df_여행사 = df_raw[df_raw['gb2']=='여행사']
grouped = df_여행사['v1m'].groupby(df_여행사['ta_ym'])
t_여행사 = pd.DataFrame(grouped.mean())
t_여행사
```

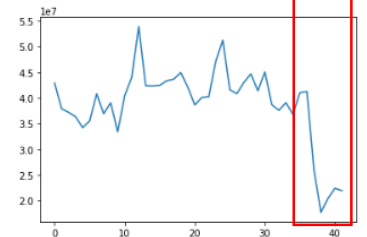
```
In [27]: df_체험 = df_raw[df_raw['gb2']=='체험']
grouped = df_체험['v1m'].groupby(df_체험['ta_ym'])
t_체험 = pd.DataFrame(grouped.mean())
t_체험
```



매출액 그래프를 통해  
코로나 이후 급격히  
줄어든 매출 확인

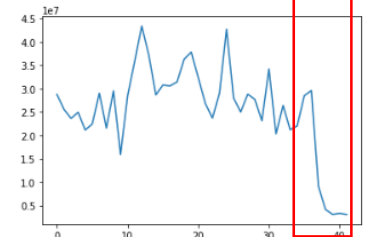
```
In [48]: data교통 = pd.DataFrame()
data교통['time'] = [i for i in range(1, 43)]
data교통['v1m'] = np.array(t_교통['v1m'])
data교통['v1m'].plot()

Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x2090e152cf8>
```



```
In [49]: data여행 = pd.DataFrame()
data여행['time'] = [i for i in range(1, 43)]
data여행['v1m'] = np.array(t_여행['v1m'])
data여행['v1m'].plot()

In [50]: <matplotlib.axes._subplots.AxesSubplot at 0x2090e20fe8>
```





### 3. 데이터 분석 – 데이터 차분 비교

#### 숙박

#### 교통

#### 관광쇼핑

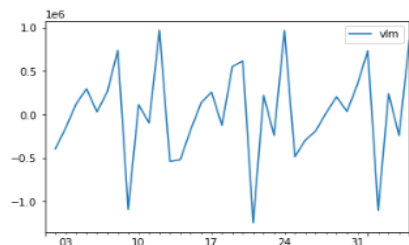
#### 여행사

1차 차분

2차 차분

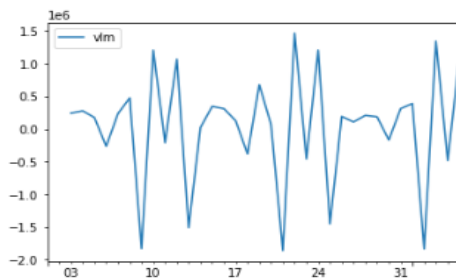
```
# 1번 차분:  $z(t) - z(t-1)$ 
# shift(n): 앞의 n번 행의 데이터를 가져옴
# df_CNY_diff1 = df_CNY_train["STD_RATE"] - df_CNY_train["STD_RATE"].shift(1)
df_숙박_diff1 = df_숙박_train - df_숙박_train.shift(1)
df_숙박_diff1.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x2070f362198>



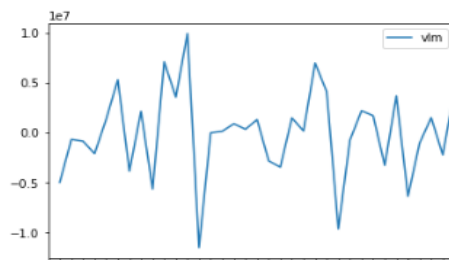
```
# 2번 차분:  $(z(t) - z(t-1)) - (z(t-1) - z(t-2)) = z(t) - z(t-2)$ 
df_숙박_diff2 = df_숙박_train - 2*(df_숙박_train.shift(1))
df_숙박_diff2.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x2070eed1a9>



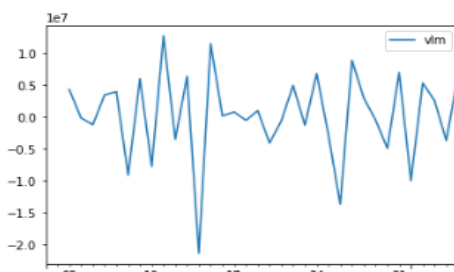
```
# 1번 차분:  $z(t) - z(t-1)$ 
# shift(n): 앞의 n번 행의 데이터를 가져옴
# df_CNY_diff1 = df_CNY_train["STD_RATE"] - df_CNY_train["STD_RATE"].shift(1)
df_교통_diff1 = df_교통_train - df_교통_train.shift(1)
df_교통_diff1.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x20710a7abe0>



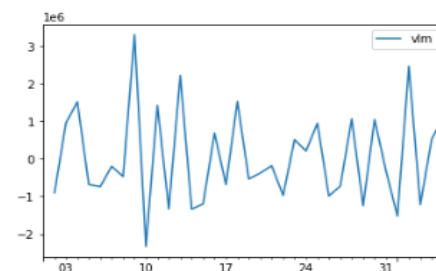
```
# 2번 차분:  $(z(t) - z(t-1)) - (z(t-1) - z(t-2)) = z(t) - z(t-2)$ 
df_교통_diff2 = df_교통_train - 2*(df_교통_train.shift(1))
df_교통_diff2.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x207106dc32f>



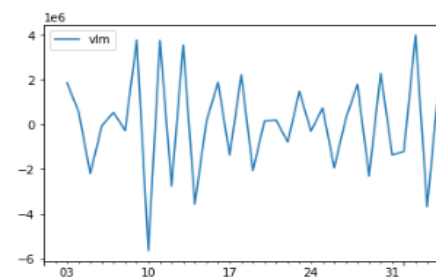
```
# 1번 차분:  $z(t) - z(t-1)$ 
# shift(n): 앞의 n번 행의 데이터를 가져옴
# df_CNY_diff1 = df_CNY_train["STD_RATE"] - df_CNY_train["STD_RATE"].shift(1)
df_관광쇼핑_diff1 = df_관광쇼핑_train - df_관광쇼핑_train.shift(1)
df_관광쇼핑_diff1.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x2070eccc3>



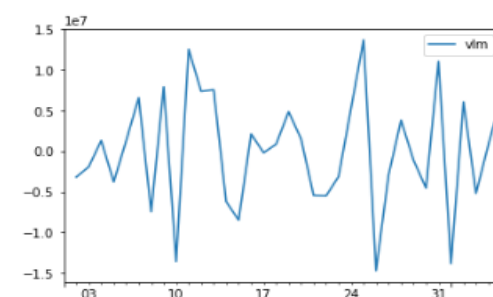
```
# 2번 차분:  $(z(t) - z(t-1)) - (z(t-1) - z(t-2)) = z(t) - z(t-2)$ 
df_관광쇼핑_diff2 = df_관광쇼핑_train - 2*(df_관광쇼핑_train.shift(1))
df_관광쇼핑_diff2.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x20710dd4ac8>



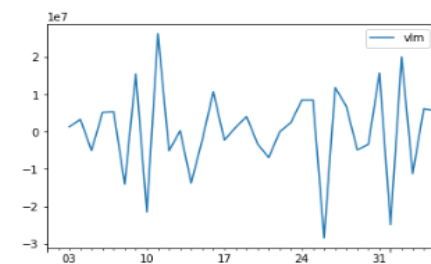
```
# 1번 차분:  $z(t) - z(t-1)$ 
# shift(n): 앞의 n번 행의 데이터를 가져옴
# df_CNY_diff1 = df_CNY_train["STD_RATE"] - df_CNY_train["STD_RATE"].shift(1)
df_여행사_diff1 = df_여행사_train - df_여행사_train.shift(1)
df_여행사_diff1.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x20710b34588>



```
# 2번 차분:  $(z(t) - z(t-1)) - (z(t-1) - z(t-2)) = z(t) - z(t-2)$ 
df_여행사_diff2 = df_여행사_train - 2*(df_여행사_train.shift(1))
df_여행사_diff2.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x2070fdb50>



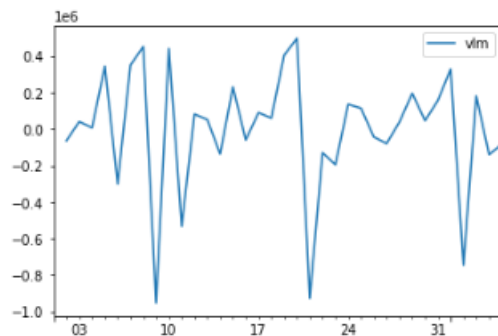
계절 변수와 시기 별 변수들을 고려하여 연도 기준으로 차분을 하면 좋으나,  
월별 데이터로 인하여 1차 차분으로 정상성을 만족시키도록 한다.

### 3. 데이터 분석 – ARIMA분석

#### 체험

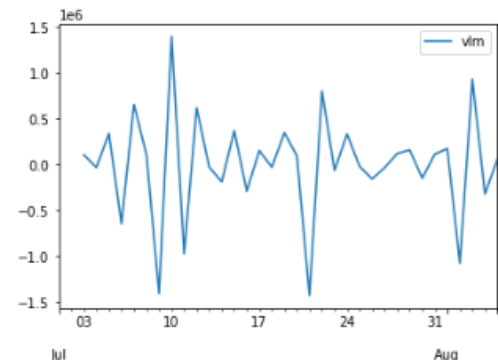
```
# 1번 차분:  $z(t) - z(t-1)$ 
# shift(n): 앞의 n번 패 행의 데이터를 가져옴
# df_CNY_diff1 = df_CNY_train["STD_RATE"] - df_CNY_train["STD_RATE"].shift(1)
df_체험_diff1 = df_체험_train - df_체험_train.shift(1)
df_체험_diff1.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x20a972424>

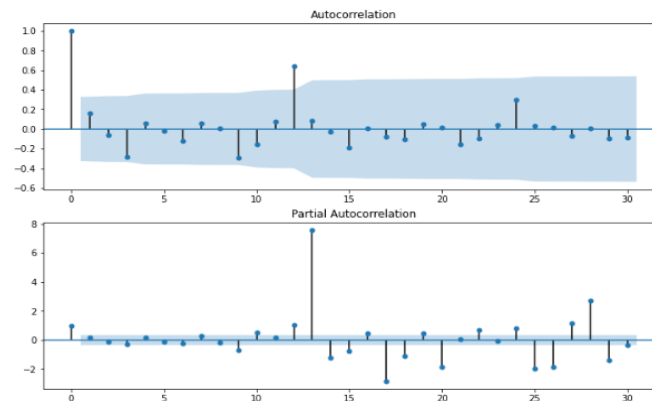


```
# 2번 차분:  $(z(t) - z(t-1)) - (z(t-1) - z(t-2)) = z(t) - 2z(t-1) + z(t-2)$ 
df_체험_diff2 = df_체험_train - 2*(df_체험_train.shift(1)) + df_체험_train.shift(2)
df_체험_diff2.plot()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x2070f55c4f>

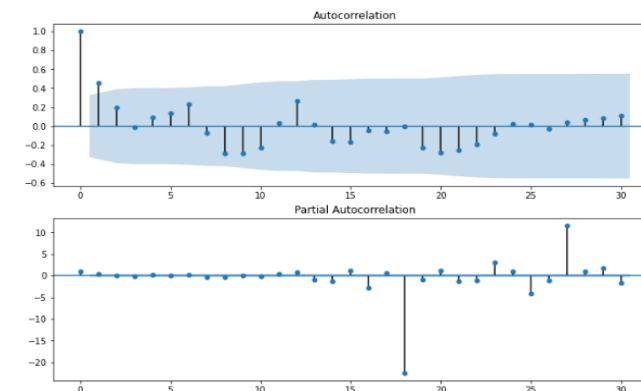


#### 숙박



ARIMA Model Results						
=====						
Dep. Variable:	D2.vlm	No. Observations:	34			
Model:	ARIMA(1, 2, 0)	Log Likelihood	-505.686			
Method:	css-mle	S.D. of innovations	692043.034			
Date:	Tue, 11 Aug 2020	AIC	1017.372			
Time:	01:03:21	BIC	1021.951			
Sample:	07-03-2017	HQIC	1018.934			
	- 08-05-2017					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
const	3.682e+04	7.49e+04	0.492	0.623	-1.1e+05	1.84e+05
ar.L1.D2.vlm	-0.6077	0.135	-4.511	0.000	-0.872	-0.344
Roots						
-----						
	Real	Imaginary	Modulus	Frequency		
-----						
AR.1	-1.6454	+0.0000j	1.6454	0.5000		
-----						

#### 교통



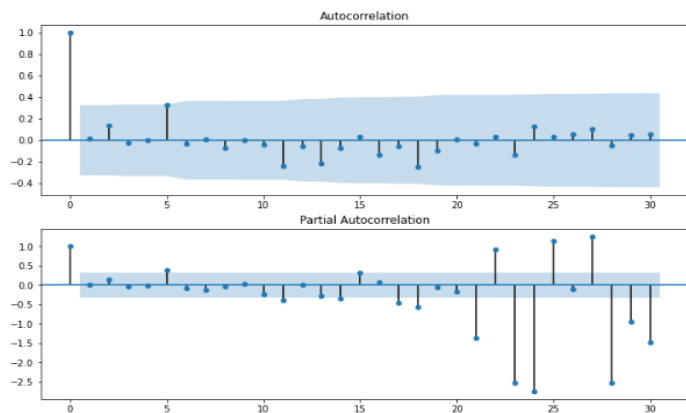
ARIMA Model Results						
=====						
Dep. Variable:	D2.vlm	No. Observations:	34			
Model:	ARIMA(1, 2, 0)	Log Likelihood	-577.483			
Method:	css-mle	S.D. of innovations	5723434.856			
Date:	Tue, 11 Aug 2020	AIC	1160.965			
Time:	01:04:04	BIC	1165.544			
Sample:	07-03-2017	HQIC	1162.527			
	- 08-05-2017					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
const	2.7e+05	6.34e+05	0.426	0.670	-9.72e+05	1.51e+06
ar.L1.D2.vlm	-0.5696	0.139	-4.106	0.000	-0.841	-0.296
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		
AR.1	-1.7557	+0.0000j	1.7557	0.5000		

1차 차분

2차 차분

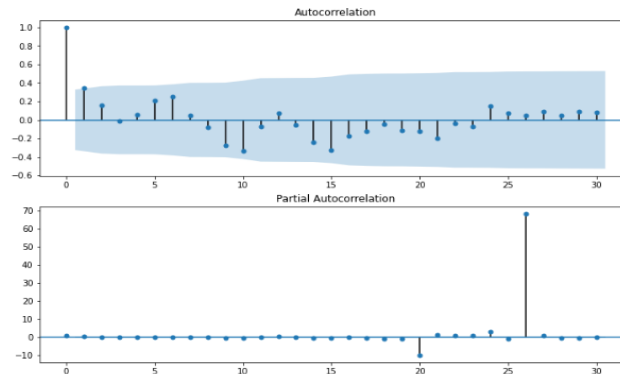
### 3. 데이터 분석 – ARIMA분석

#### 관광쇼핑



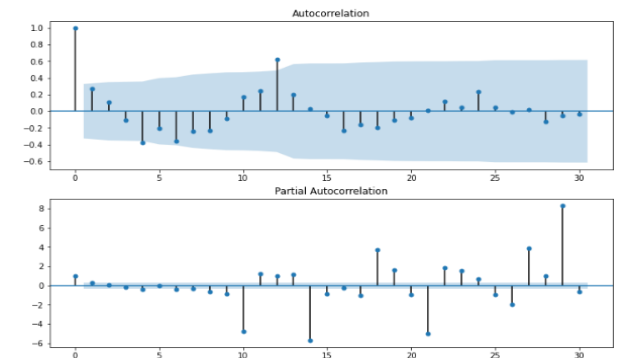
ARIMA Model Results						
Dep. Variable:	D2.vlm	No. Observations:	34			
Model:	ARIMA(1, 2, 0)	Log Likelihood	-532.854			
Method:	css-mle	S.D. of innovations	1532497.802			
Date:	Tue, 11 Aug 2020	AIC	1071.708			
Time:	01:04:25	BIC	1076.287			
Sample:	07-03-2017 - 08-05-2017	HQIC	1073.270			
	coef	std err	z	P> z	[0.025	0.975]
const	5.894e+04	1.55e+05	0.381	0.703	-2.45e+05	3.62e+05
ar.L1.D2.vlm	-0.7220	0.113	-6.410	0.000	-0.943	-0.501
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-1.3850	+0.0000j	1.3850	0.5000		

#### 여행사



ARIMA Model Results						
Dep. Variable:	D2.vlm	No. Observations:	34			
Model:	ARIMA(1, 2, 0)	Log Likelihood	-593.508			
Method:	css-mle	S.D. of innovations	9158884.418			
Date:	Tue, 11 Aug 2020	AIC	1193.015			
Time:	01:04:44	BIC	1197.595			
Sample:	07-03-2017 - 08-05-2017	HQIC	1194.577			
	coef	std err	z	P> z	[0.025	0.975]
const	2.834e+05	9.85e+05	0.288	0.774	-1.65e+06	2.21e+06
ar.L1.D2.vlm	-0.6132	0.130	-4.723	0.000	-0.868	-0.359
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-1.6307	+0.0000j	1.6307	0.5000		

#### 체험



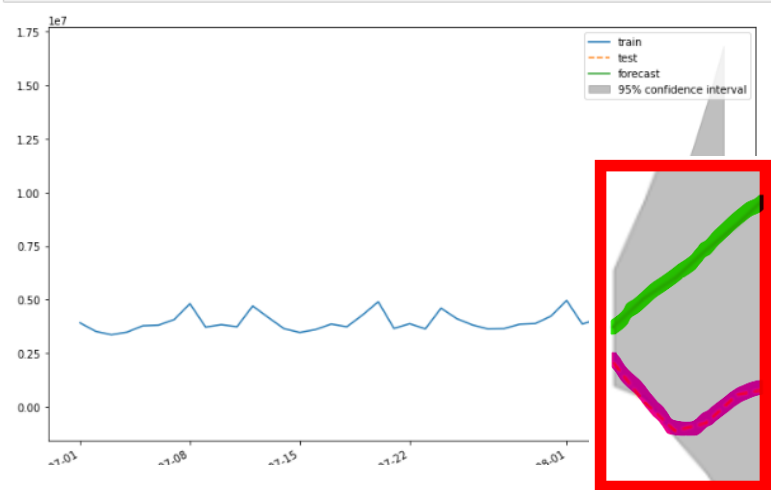
ARIMA Model Results						
Dep. Variable:	D2.vlm	No. Observations:	34			
Model:	ARIMA(1, 2, 0)	Log Likelihood	-490.261			
Method:	css-mle	S.D. of innovations	439176.955			
Date:	Tue, 11 Aug 2020	AIC	986.523			
Time:	01:05:24	BIC	991.102			
Sample:	07-03-2017 - 08-05-2017	HQIC	988.084			
	coef	std err	z	P> z	[0.025	0.975]
const	-562.5004	4.64e+04	-0.012	0.990	-9.15e+04	9.03e+04
ar.L1.D2.vlm	-0.6433	0.125	-5.160	0.000	-0.888	-0.399
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-1.5544	+0.0000j	1.5544	0.5000		

ACF가 감소하며, PACF가 2시점에서 절단면을 갖는다면, 1시점 전의 자료가 현재에 영향을 미침 → AR(1) 모델

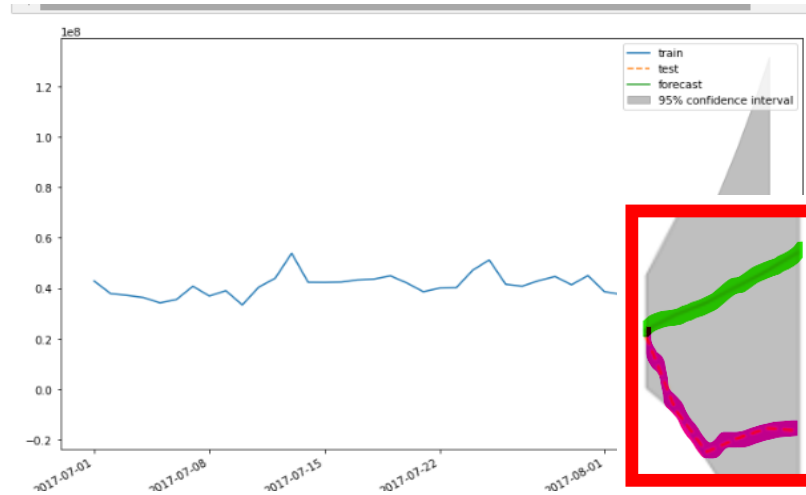


### 3. 데이터 분석 – 신한 카드 데이터 분석

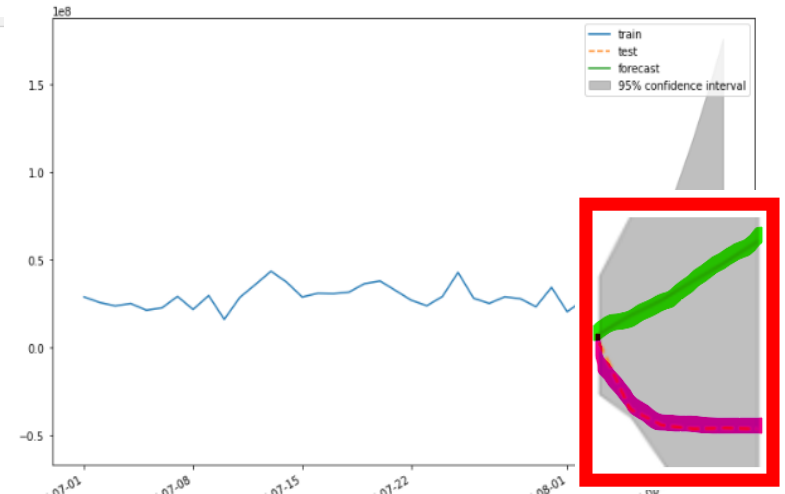
숙박



교통



여행사

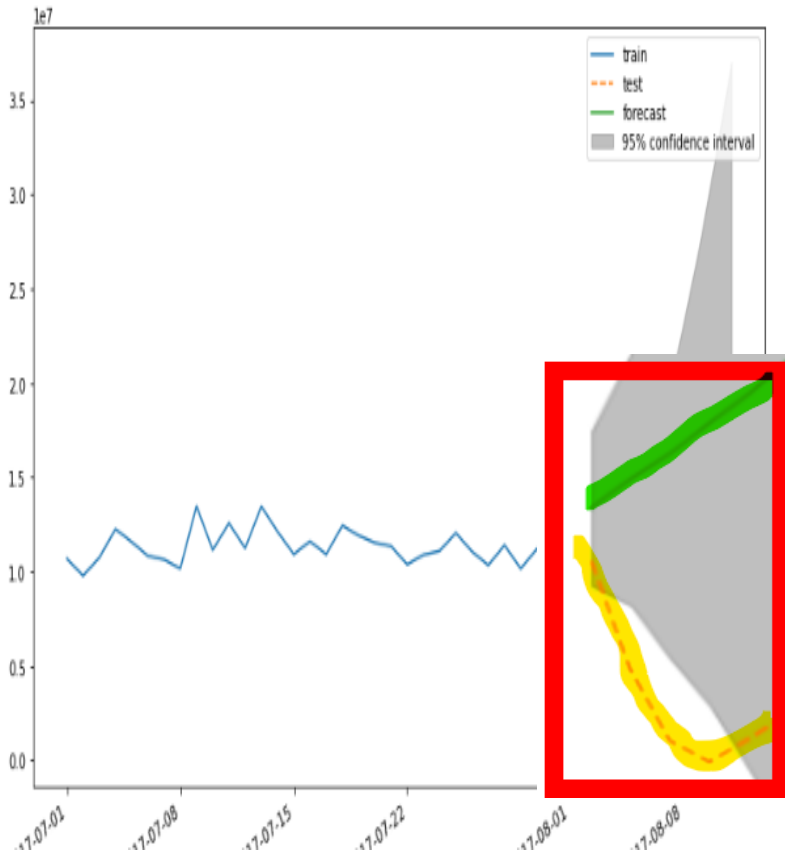


2020년 1월 이후를 Test Data로 두어,  
기존의 예상매출액보다 얼마나 여행산업에서의  
소비가 침체 되었는지를 확인할 수 있다.

세 변수 모두 상승하는 **예측치**에 비하여  
**급격하게 하락**하는 실제 TEST 매출액을 보이고 있다.

### 3. 데이터 분석 – 신한 카드 데이터 분석

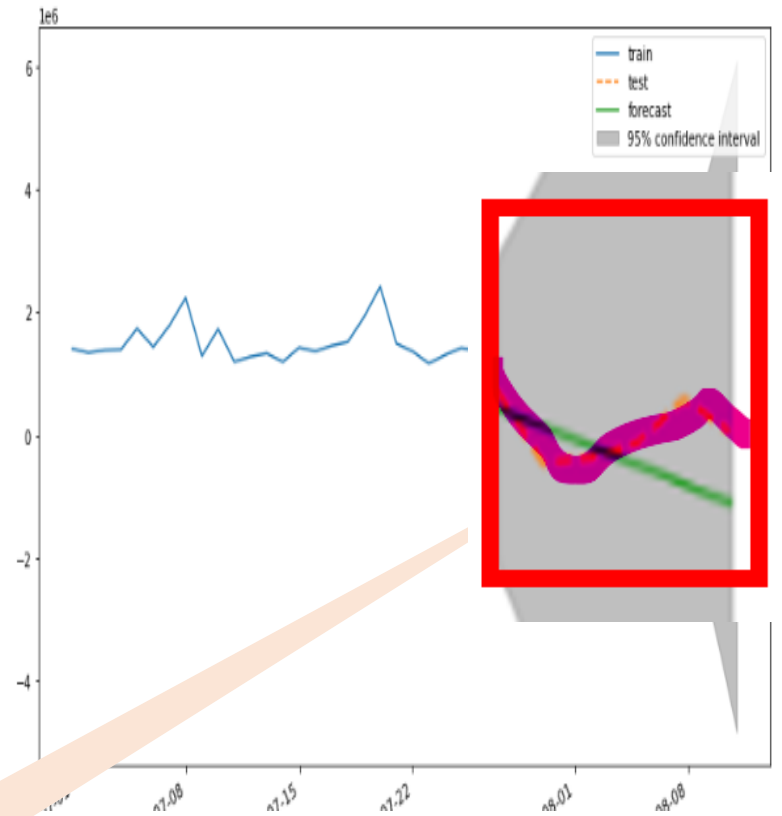
관광쇼핑



특히 **관광쇼핑** 분야에서  
코로나의 여파가 가장 크게 나타  
났다. 관측 예상치 95%를 넘어  
서는 수치를 보이며 코로나 이후  
매출의 급감을 나타낸다.

반면, **체험** 분야에서는 두드러  
지는 하락이 매출액의 하락이 발  
생하지 않았다. 이후에는 기저 효  
과일 수 있지만, 예상치보다 높  
은 매출액을 보이기도 한다.

체험

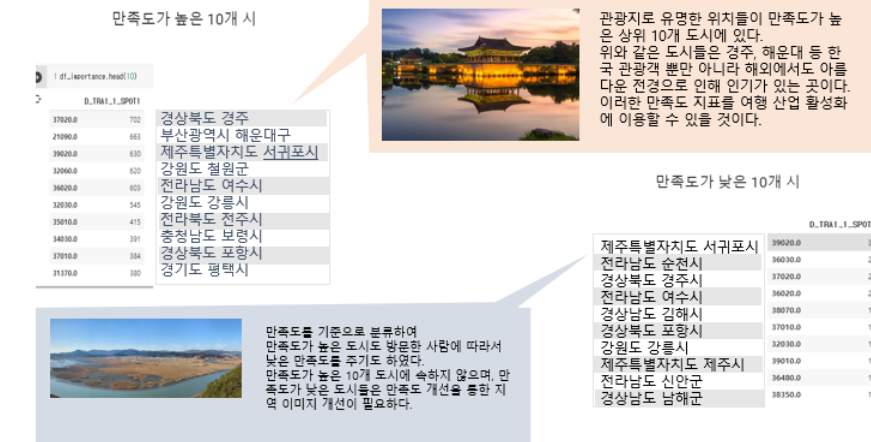
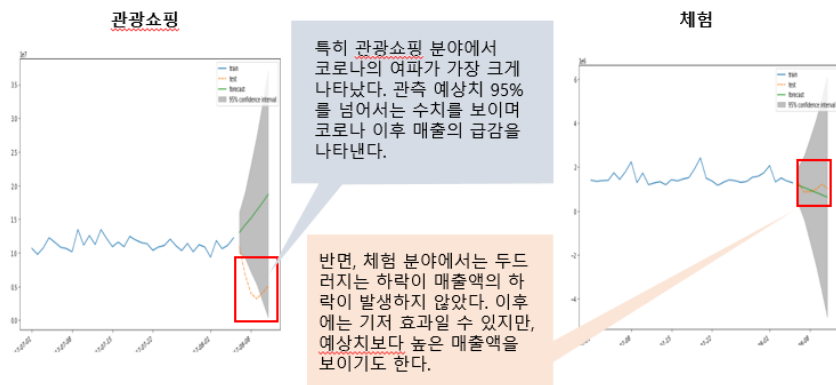


## 5. 결론 및 시사점



### 결과 정리 및 시사점

- 지금까지의 분석을 통하여 여행 산업에서 **코로나의 영향**과 이전의 **국민의 여행 수요**를 확인해 볼 수 있었다.



- 여행의 만족도에 가장 큰 영향을 주는 요인으로는 **장소**와 **가격**이 있다.
- 코로나로 인해 가장 매출액에 큰 변화가 있는 여행산업은 **관광쇼핑**이고, 가장 영향이 적은 산업은 **체험**이다.



## 5. 결론 및 시사점



### 결과 정리 및 시사점

1. 여행의 만족도에 가장 큰 영향을 주는 요인으로는 **장소**와 **가격**이 있다.

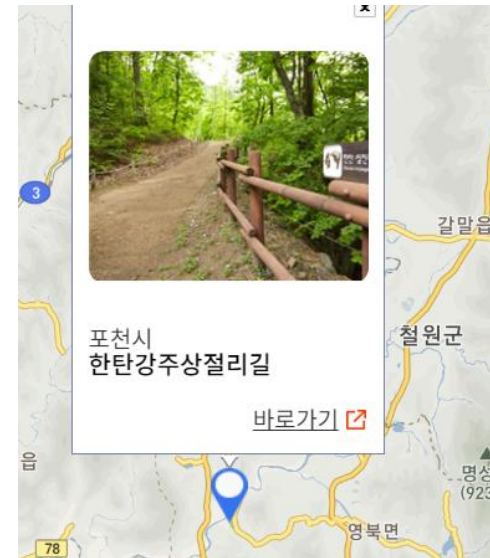
코로나 상황에서 문화 관광 산업의 활성화를 위해서는 안전하게 여가생활 하면서도, 만족도가 높은 여행을 유도해야 한다.

여행지  
장산 (부산 국가지질공원)  
부산 해운대구 | 내 위치 OFF



만족도가 높은 부산의 비대면 여행지 장산

또한 기존의 사람이 많던 관광지에서 가깝게 이동해 갈 수 있는 관광지를 선별하여 홍보할 경우, 만족도를 높이며, 타 지역과 상호 발전할 수 있다.



만족도가 높았던 강원도 철원군 근처의 비대면 여행지한탄강 주상절리길

## 5. 결론 및 시사점



### 결과 정리 및 시사점

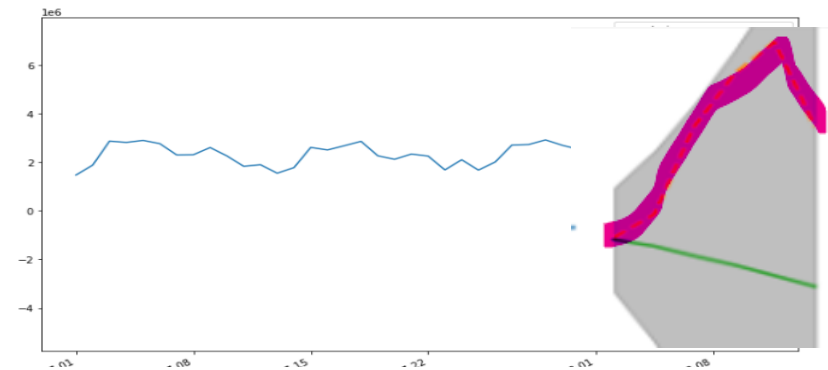
2. 코로나로 인해 가장 매출액에 큰 변화가 있는 여행산업은 **관광 쇼핑**이고, 가장 영향이 적은 산업은 **체험**이다.

이와 동시에 사람들이 안전하게 느끼는 체험 활동을 비대면 여행지에서 진행할 경우 더 사람들이 안전하다고 생각하며 여행을 즐길 수 있을 것이라 예측된다.

대형 여행사를 통한 관광을 즐기며 쇼핑을 하기보다는 캠핑과 스포츠에 대한 수요가 높아지고 있다.

비대면을 위한  
개인 여행( 차박, 캠핑, 자전거 등 )에 체험 활동과  
기존의 만족도를 기반으로 한  
국내 비대면 관광지 활성화가 필요하다.

```
# 예측 시작, 종료 날짜
start_time = df_자전거_test.index.min() # 2018-03-23
end_time = df_자전거_test.index.max() # 2018-03-31
fig, ax = plt.subplots(figsize=(12, 8))
plt.plot(df_자전거_train, linestyle='--', label='train')
plt.plot(df_자전거_test, linestyle='--', label='test')
# 예측 그래프 생성 함수, start: 예측 시작, end: 예측 종료, plot_insample: 가지고 있는 데이터 중
fig = ts_result_자전거.plot_predict(start=start_time, end=end_time, ax=ax, plot_insample=False)
plt.show()
```



시계열 분석으로 확인되는  
늘어난 자전거 소비



산업 > 자동차

### [Car&Fun]'언택트' 시대...나만의 차박으로 '힐링'

입력 2020-05-17 10:00:01 수정 2020.05.17 18:44:53 박한신 기자



The background image is a traditional Korean landscape. It features a dark, silhouetted mountain range in the background. In the middle ground, there is a traditional Korean pavilion (Jeongja) with a tiled roof and ornate details, situated on a grassy area. To the left, there is a pond with a stone bridge and a stone wall. The foreground is a dark, silhouetted area, possibly a path or a field. The overall scene is captured in a dark, moody style with a semi-transparent dark overlay in the center.

# Thank You

---

DATA89