



포스트 코로나

(Post COVID19)시대,  
20대의 소비 생활

-제안 보고서-

제출 문서	다변량 통계학 학기말 과제
담당 교수	김규성 교수님
제출 기한	2020.11.23
실 제출일	2020.11.23
학과/학년	통계학과 4학년
학번	2014580002
이름	김광클

## 1. 서론

- 1.1 연구목적
- 1.2 문헌 연구
- 1.3 데이터 설명
- 1.4 분석 방법

## 2. 예상되는 결과

# 1. 서론

## 1.1 연구 목적

2019년 말부터 현재 2020년까지 COVID19 바이러스(이하 코로나)는 세계인들의 생활양식을 바꾸어 놓았다. 세계는 현재 코로나 이전과 코로나 이후(포스트 코로나시대)로 나누어져서 구분 된다고 해도 과언이 아닐 정도로 코로나가 가져온 변화는 극심하다.

코로나의 많은 문제들 중에서도 가장 큰 문제점 중 하나는 바로 경제활동의 변화이다. 사람들이 밖에 나가는 것을 꺼리며, 대면과 만남에 대한 두려움과 함께 각각의 소비패턴이 저하되거나 비대면의 식으로 바뀌어 나가는 사회 현상을 목격 할 수 있었다.

대한민국 또한 코로나에 대해 다양한 정책들과 예방책들을 내놓았으며 그에 따른 사람들의 소비변화가 나타났다. 그 중에서도 우리는 20대의 소비행동에 집중해 보기로 하였다.

20대의 경우 다양한 계층(학생, 직장인, 무직, 사업가, 군인 등)을 가짐으로써 개개인의 소비행동이 특정 나이대보다 개성이 있을 가능성이 있다고 판단하였고, 무엇보다도, 외부로의 사회적 만남(대면)이 그 어떠한 나이대보다 활발한 이들이 소비행동과 활동들이 포스트코로나사회에서 어떻게 바뀌었는지를 찾고자 한다.

## 1.2 문헌 연구

<포스트 코로나시대: 나이, 성별, 카테고리 별 매출 변화 분석>

<https://dacon.io/competitions/official/235618/codeshare/1525?page=1&dtype=recent&ptype=pub>

<포스트코로나 시대에서 발견한 소비현상과 떠오르는 물음, 그리고 그것의 답>

<https://dacon.io/competitions/official/235618/codeshare/1448?page=2&dtype=vote&ptype=pub>

<포스트 코로나 소비와 흥미의 변화는??>

<https://dacon.io/competitions/official/235618/codeshare/1419?page=1&dtype=vote&ptype=pub>

### 1.3 데이터 설명

총 8개의 데이터셋을 사용할 예정이며, 각각의 데이터별 칼럼에 대한 설명을 표로 정리 하였다.

Key 데이터							
-----							
fpopl.csv(지역별 성별, 연령별 인구수)							
base_ymd 기준날짜	tmzon_se_code 기준시간	sexdstn_se_code 성별	agrde_se_code 연령대별	adstrd_code 행정통코드별	popltn_cascnt 인구수		
adstrd_master.csv(지역명, 지역코드)							
adstrd_code 행정통코드	adstrd_nm 행정동명	brtc_nm 시/도 명	signgu_nm 시/군/구 명				
-----							
코로나 확진자 관련 데이터							
-----							
Time.csv (코로나 확진 일자별 '누적' 데이터, 2020.1.20 ~ 2020.06.30)							
date 날짜	time 시간	test 검사자수	negative 음성자수	confirmed 확진자수	released 격리해제	deceased 격리해제	
TimeAge.csv (코로나 확진자 일자별 나이대별 데이터, 2020.03.02 ~ 2020.06.30)							
date 날짜	time 시간	age 나이대	confirmed 확진자수	deceased 사망자수			
PatientInfo_20200717.csv (코로나 확진자 개인별 정보)							
patient_id 환자식별번호	sex 성별	age 나이대	country 국가	province 시/도	city 시/군/구		
infection_case 감염경로	infected_by 감염자	contact_number 접촉사람수	symptom_onset_date 증상발현날짜	confirmed_date 확진날짜	released_date 격리해제날짜	deceased_date 사망날짜	state 현상황(격리해제, 사망, 격리)
-----							
분석이용 데이터							
-----							
Policy.csv (정부정책 및 사건)							
policy_id 사건번호	country 국가(Korea로 통일)	type 사건별대분류	gov_policy 사건별소분류	detail 구체적사건명	start_date 사건시작일	end_date 사건종료일	
index.csv (상품별(나이별,성별,지역별) 구매 목록, 성장지수 데이터 2019.1 ~ 2020.5)							
period 년/월	catl 상품대분류	catm 상품소분류	age 나이대	gender 성별	sido 시/도	sigungu 시/군/구	cgi 카테고리성장지수
card_20200717.csv (2020.01.04 ~ 2020.03.05 개인별 카드 사용 내역 데이터)							
receipt_dttm 카드사용날짜	adstrd_code 행정통코드	adstrd_nm 행정동이름	mrhst_induty_cl_code 업종코드	mrhst_induty_cl_nm 업종명	selng_cascnt 매출건수	salamt 매출금액	
-----							
추가이용 가능 데이터							
-----							
delivery.csv(배달관련 데이터, 2020.01.01 ~ 2020.02.08, 다수의 캠퍼스를 보유하고 있다.)							

## 1.4 분석 방법

(0)

데이터를 축소하겠다.(지역의 한정)

인구밀집도가 높고, 분석의 타깃층으로 하는 20대가 많이 몰려있으며, 경제활동도 가장 활발하고, 데이터가 많은 '서울'의 데이터만을 중심으로 분석하고자 한다.

(1)

포스트 코로나시대를 정의하겠다. (코로나 시대 이전 vs 코로나 시대 이후)

일별 확진자수 데이터(Time.csv)를 이용하여 '탐색적데이터분석(EDA)'를 실시하고 기준(합리적인 이유가 있어야 할 것)을 잡아서 포스트코로나시대의 정확한 정의(시기)를 내리도록 하겠다.

(2)

포스트 코로나 이전과 이후의 연령별 소비 패턴의 변화 분석을 하겠다.

'다변량 분산분석'을 통해 연령별 (코로나 이전 vs 코로나 이후)의 소비량의 변화가 있었는지(통계적으로 유의미하게 변화였다.) 분석해보도록 한다.

(3)

20대의 소비품목(대분류별)을 분석(EDA이용가능)한 이후 (코로나 이전 vs 코로나 이후)의 변화를 관찰한다.

(4) (3)의 결과에서 전체 소비량의 분산을 설명하는데 있어 치명적인 소비품목들(대분류목록중)만을 뽑아내어('PCA: 주성분분석 사용') 분석을 실시하도록 하겠다.

(5) 위의 뽑힌 변수들(대분류)중 소분류를 더 구체적으로 파악하여 20대의 소비 패턴에서 어떠한 '인사이트'를 얻을 수 있는지 분석해보도록 한다.

## 2. 예상되는 결과

(1)

1.4 - (1)에서 EDA를 실시한 결과 특정 시기에 확진자의 증가율이 급격하게 늘어난 날짜(특정사건으로 인하여: 이태원클럽, 신천지 등)가 있을 것으로 예상되며, 추후 나온 정부정책들을 함께 고려하여 시각화된 시계열 그래프를 통하여, 포스트코로나 시기를 결정할 수 있을 것으로 예상된다.

(2)

1.4 - (2)에서 비 20대의 경우에는 소비량이 줄어들었을 것이라는 유의미한 통계적 차이가 나올 것이라고 생각된다. 그러나 20대에서는 코로나에도 불구하고 소비량이 크게 줄지 않을 것이라는 결과를 예상한다.

(3)

1.4 - (3),(4)에서 20대의 소비량과 관계없이 바뀐 생활양식으로 인하여 20대의 소비하는 품목이 분명히 바뀐점이 있을 것이라고 생각한다. 분석을 통해서 코로나 이전과 이후로 비교하였을 때, 20대의 소비패턴에서 어떠한 품목들이 가장 크게 변동하였는지 알 수 있을 것이라고 생각한다.

(4)

1.4 - (5)에서 결과를 통해서 사회적 인사이트를 찾아낼 수 있도록 하고, 그 인사이트로부터 도출 될 수 있는 또 다른 분석들이 나온다면, 추가적으로 해보고자 한다.

## ※참고 문헌

데이터 출처: <https://dacon.io/competitions/official/235618/data>