

「코로나-19 사태에 따른 서울특별시 내 카페 동향」

최종보고서

2021. 12. 9

김 서 영

졸업작품최종보고서

학 부 / 전 공	데이터사이언스전공	학 번	201804194
성 명	김 서 영	졸업 예정 일	22 년 2 월
구 분	단 독 <input checked="" type="checkbox"/> 공 동 <input type="checkbox"/> ()인		
휴 대 전 화	01086229636	E-mail	rnddlrhddl@gmail.com
작 품 제 목	코로나-19 사태에 따른 서울특별시 내 카페 동향		
작 품 개 요	<p>코로나-19는 2019년 12월 중국에서 처음 발생한 후 2020년 1월 한국에서도 첫 확진자가 발생하였다. 이에 정부는 감염 확산 방지를 위하여 음식점으로 취급되는 가게에 단계별로 세분화된 방역 수칙을 지정하였다. 하지만 사태가 약 2년 동안 지속됨에 있어 폐업하는 음식점들도 많아졌다. 이 중 서울시 내 '카페'를 중심으로 폐업 요인에 대한 분석을 진행하였다.</p> <p>우선 카페 데이터를 서울시 음식점 인허가 정보를 통해 수집할 수 있었다. 다른 유형의 음식점 또한 포함되어 있는 데이터로 카페 데이터만 추출하는 전처리를 거쳤다. 카페의 폐업에 영향을 미칠 수 있는 데이터인 카페 주변(1km) 지하철역, 공공기관, 관광명소, 대학교에 대한 정보를 카카오 API서비스를 통하여 수집하였다. 서울시 주민등록인구 통계를 활용하여 카페 주변(1km)의 유동인구 정보를 파악할 수 있게 하였다.</p> <p>코드 GitHub주소 https://github.com/seoyoung047/covid-research-about-cafe</p>		
위와 같은 내용으로 20 학년도 전기/후기 졸업작품 최종보고서를 제출합니다.			
2021 년 12 월 9 일			
제 출 자 : 김서영 (인) 			
산 업 데 이 터 사 이 언 스 학 부			

※ 아래와 같은 내용의 졸업작품 최종보고서를 작성하여 제출하시오.

- 데이터 분석 기획 전 프로세스 내용 포함 (문제 사항 보완 결과 포함)
- 작품 제목, 필요성, 작품 개요(각 구성 요소별 block diagram 수준 기술), 분석 기술(플랫폼, 사용 프로그래밍 언어 및 기타 도구), 필요 장비/소프트웨어, 제작 과정, 작품 해석 등

※ 2인 이상의 공동 작품은 졸업작품 결과보고서를 각자의 담당 부분을 중심으로 기술하여 개별적으로 제출하되, 작품 전체에 대한 요약본을 첨부하시오.

승인여부	가()	부()	가()	부()	가()	부()	가()	부()
지도교수	(인)		(인)		(인)		(인)	

보 완 사 항 확 인 서

학 부 / 전 공			
성 명		졸 업 예 정 일	년 월
구 분	단 독 <input type="checkbox"/> 공 동 <input type="checkbox"/> ()인		
휴 대 전 화		E-mail	
작 품 제 목			
요 구 사 항			
최 종 결 과			

위와 같이 졸업작품을 최종 보완하였음을 확인합니다.

년 월 일
 교 수 : (인)
 제 출 자 : (인)

산 업 데 이 터 사 이 언 스 학 부

목 차

제 I 장 서론

- 1 배경 및 필요성 2
- 2 활용 방안 2

제 II 장 분석 기획

- 1 비즈니스 이해 및 범위 설정 3
- 2 프로젝트 정의 및 계획 설정 3

제 III 장 데이터 준비

- 1 필요 데이터 정의 5
- 2 데이터 수집 및 전처리 6

제 IV 장 데이터 분석

- 1 분석용 데이터 준비 11
- 2 탐색적 분석 12
- 3 모델링 17
- 4 모델 평가 18

제 V 장 평가 및 전개

- 1 프로젝트 평가 및 향후 계획 20

- 참고문헌 21

I 서론

1. 배경 및 필요성

가. 배경

- 코로나-19는 2019년 12월 중국 우한에서 처음 발생한 후 2020년 1월 한국에서도 첫 확진자가 발생하였다. 이 후 한국 정부는 마스크 착용과 사회적 거리두기를 실행하며 감염의 유입 및 확산에 대응하였다. 초기 3단계였던 사회적 거리두기 단계는 대유행을 거치며 현재는 총 5단계로 세분화 되었다.
- 특히 유동인구와 인구 밀집도가 높은 서울시는 확진자 현황에 따라 사회적 거리두기를 조정하며 코로나-19에 대한 방역을 시도하였다.

나. 필요성

- 사회적 거리두기는 가게의 형태별로 다른 방역 수칙을 지정하고 있다. 특히 일반 음식점과 휴게음식점 등은 마스크를 벗고 음식 섭취를 하며 대화가 오고가는 중 감염이 많이 발생하였다. 정부는 이에 대한 감염 방지를 위하여 음식점으로 취급되는 가게에 단계별로 세분화된 방역 수칙을 지정하였다.
- 하지만 높은 단계의 사회적 거리두기 시행 지속으로 인하여 많은 소상공인 및 자영업자가 영업에 어려움을 겪고 있다. 사태가 1년 이상 지속되며 폐업하는 음식점들도 속출하고 있다. 이 중 서울특별시 내 '카페'를 중심으로 폐업 요인에 대해 분석하고자 한다.

2. 활용 방안

- 현재 코로나-19 이후로 관련 분석이 활발하게 진행되고 있다. 그 중에서 이 프로젝트는 코로나-19 사태로 타격을 많이 받는 소상공인에게 필요한 정보, 즉 현재의 가게를 유지할 수 있는 방법을 제공할 수 있다. 현재 카페 개업 후 2년 내에 폐업하는 가게 들이 많다. 현 상황에서 카페 개업을 고민하는 창업자에게는 개업 시 참고해야하는 요인들의 분석으로서 도움이 될 수 있다.
- 또한 거시적으로 정부의 사회적 거리두기에 관련한 정책 구성에 도움이 될 수 있다. 악화 된 상황으로 인하여 거리두기 단계가 높아짐에 따라 아예 영업을 하지 못하는 가게들과 영업을 하여도 장사가 되지 않는 가게들이 많다. 무조건 적인 영업 제한이 아닌 세부적인 정책 구성에 도움이 될 수 있을 것이라 생각된다.

Ⅱ 분석 기획

1. 비즈니스 이해 및 범위 설정

가. 비즈니스 이해

- ☐ 이 분석에서 말하는 '카페'란 커피 전문점으로서 프랜차이즈 카페와 개인 카페를 모두 포함한다.
- ☐ 커피전문점은 커피를 포함한 다양한 음료 판매를 목적으로 하며 이용 고객은 매장 내 좌석을 이용하기도 한다.
- ☐ 카페에는 앉아서 음료와 디저트를 먹으며 대화할 수 있는 넓은 매장과 오로지 음료 테이크-아웃을 위한 매장 등으로 구분된다.

나. 범위 설정

- ☐ 분석의 범위는 서울특별시 내, 현재까지 운영 중이거나 2017년 이후 2021년 이전 폐업한 커피전문점을 대상으로 한다.
- ☐ 대상 범위를 서울특별시 내로 설정한 이유는 서울은 수도로서 인구와 카페의 큰 밀집도를 지니고 있어 코로나-19로 인한 타격이 컸기 때문이다.
- ☐ 코로나 사태의 전과 후를 비교하기 위하여 발생 약 2년 전인 2017년부터 현재 2021년까지 비교 대상으로 지정한다.

2. 프로젝트 정의 및 계획 설정

가. 프로젝트 정의

이 프로젝트는 코로나-19 사태로 인하여 폐업하게 된 가게와 계속 영업 중인 가게의 차이점을 설명하며 폐업에 영향을 준 요인에 대하여 설명한다.

나. 계획 설정

분석 요인에 해당 가게 건물의 공시지가, 주변 상권, 교통시설(지하철), 학교 등 다양한 요인을 통해 분석을 할 예정이다.

분석에서 필요한 데이터는 서울특별시에서 제공하는 '서울 열린데이터광장'에서 '서울특별시 휴게음식점 인허가 정보', '일반음식점 인허가 정보'를 활용하여 서울시 내 카페의 개·폐업 정보, 주소, 소재지 면적 등의 데이터를 구할 예정이다. 또한 '서울특별시 주민등록인구 통계'를 활용하여 유동인구 파악, '서울시 개별 공시지가 정보'를 활용하여 공시지가 데이터를 구할 예정이다. 또한 음식점 인허가 정보 내의 x , y 좌표 변환, 주변 상권, 교통시설 파악을 위하여 Kakao API 서비스를 활용할 예정이다.

III 데이터 준비

1. 필요 데이터 정의

가. '카페' 데이터

서울특별시 내 커피전문점을 말한다. 서울 열린 데이터 광장의 '일반음식점 인허가 정보'와 '휴게 음식점 인허가 정보'에서 '카페'로 분류되는 데이터만 모아 정리한다. 데이터에는 카페의 관리번호, 인허가 일자(개업일), 영업상태코드, 폐업 일자, 소재지 면적, 상호명, 위치(x, y 좌표/지번 주소/ 도로명 주소) 등을 포함한다.

나. 주변 상권 / 교통시설

카페 주변의 상권(대학교, 공공시설, 관광명소)와 교통시설(지하철역)에 대한 정보로 카카오API 서비스를 사용하여 수집한다. 카페의 위치(x, y) 좌표를 변환, 삽입 방식으로 정보를 불러오며 이 데이터에는 각 키워드의 카페 주변(반경 1km 내) 개수와 지하철역 명 등을 포함한다.

다. 공시지가

서울 열린 데이터 광장에서 제공하는 데이터로 '서울시 개별공시지가 정보'이다. API 주소에 카페의 구/동/본번/부번 정보를 삽입하여 출력하는 형태로 본번과 부번 정보는 카카오API 서비스를 통해 수집하였다. 이는 카페 건물의 대략적인 임대료, 월세 등을 파악하기 위한 데이터이다.

라. 유동인구

서울 열린 데이터 광장에서 '서울특별시 주민등록인구 통계'를 제공한다. 구/동 별 인구의 나이 대에 따라 나뉘어 제공되며 이를 통해 대략적인 유동인구를 파악할 수 있다.

마. 경쟁업체

기준 카페 비교 경쟁 업체 데이터를 전처리를 통해 구한다. 좌표로 거리를 구하여 주변 1km내 가게들 비교한다. (개업 : 운영 중인 가게, 폐업 : 폐업한 가게 / 2020년 2월 기준)

- 개업-개업(op_op), 개업-개업/개업날짜 12개월 내(op_op_12), 개업-개업/개업날짜 6개월 내(op_op_6), 개업-개업/개업날짜 3개월 내(op_op_3),

- 개업-폐업(op_cl), 개업-폐업/개업날짜 12개월 내(op_cl_12), 개업-폐업/개업날짜 6개월 내(op_cl_6), 개업-폐업/개업날짜 3개월 내(op_cl_3),
- 폐업-개업(cl_op), 폐업-개업/폐업날짜 12개월 내(cl_op_12), 폐업-개업/폐업날짜 6개월 내(cl_op_6), 폐업-개업/폐업날짜 3개월 내(cl_op_3),
- 폐업-폐업(cl_cl), 폐업-폐업/폐업날짜 12개월 내(cl_cl_12), 폐업-폐업/폐업날짜 6개월 내(cl_cl_6), 폐업-폐업/폐업날짜 3개월 내(cl_cl_3)

라. 확진자 데이터

‘서울 열린데이터 광장’에서 제공하는 데이터로 csv 파일로 제공된다. 서울특별시 구별로 매일의 누적, 추가 확진자로 데이터가 제공된다.

2. 데이터 수집 및 전처리

가. ‘카페’ 데이터 수집

우선 서울 열린 데이터 광장의 ‘일반음식점 인허가 정보’의 API를 통해 서울시 내 음식점 데이터를 불러오고 필요한 출력값만 정리하여 저장한다. ‘휴게 음식점 인허가 정보’는 이와 동일한 방식이다.

나. ‘카페’ 데이터 전처리

전처리 전의 일반음식점 데이터의 업태구분 목록이다. 분석에서 사용할 카페 데이터(7931개)와 커피숍 데이터(5개)를 추출한다.

간이주점	1	이동조리	102
감성주점	319	일반조리판매	6
경양식	46633	일식	16719
기타	28549	전통찜집	831
기타 휴게음식점	4	정종/대포집/소주방	12743
김밥(도시락)	1454	제과점영업	1
카페	7931	중국식	14193
냉면집	86	출장조리	355
다방	3	커피숍	5
도매업(유통)	1	키즈카페	90
라이브카페	234	당류(보신용)	408
통살품	2	통닭(치킨)	9582
볶어취급	108	패밀리레스토랑	176
분식	78247	패스트푸드	3927
튀폐식	2754	한식	194359
식육(숯불구이)	1340	호프/통닭	33987
식품등 수입판매업	1	회집	1148
식품소분업	1	횃집	209
식품제조가공업	1		
외국음식전문점(인도,태국등)	1517		

전처리 전의 휴게음식점 데이터의 업태구분 목록이다. 분석에서 사용할 커피숍 데이터(26772개)만 추출한다.

고속도로	9	유원지	59
공항	18	일반조리판매	18605
과자점	8681	전통찜집	340
관광호텔	80	철도역구내	436
극장	19	커피숍	26772
기타	2	키즈카페	129
기타 휴게음식점	14724	패스트푸드	7737
김밥(도시락)	1	편의점	9194
다방	17695	푸드트럭	952
단란주점	1	한식	2
덱카페	136	호프/통닭	1
룸살롱	1		
백화점	2633		
아이스크림	426		

위의 두 데이터를 병합한 후 저장한다.(cfdata)

```
In [182]: cfdata.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34708 entries, 0 to 34707
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype  
---  --
0   APVPERMYMD      34708 non-null  datetime64[ns]
1   TRDSTATEGBN     34708 non-null  int64   
2   TRDSTATENM      34708 non-null  object  
3   DCBYMD          18525 non-null  datetime64[ns]
4   SITEAREA        32668 non-null  object  
5   RDNWHLADDR      30656 non-null  object  
6   BPLCNM          34708 non-null  object  
7   UPTAENM         34708 non-null  object  
8   X               34279 non-null  float64 
9   Y               34279 non-null  float64 
dtypes: datetime64[ns](2), float64(2), int64(1), object(5)
memory usage: 2.6+ MB
```

APVPERMYMD	인허가일자
TRDSTATEGBN	영업상태코드(1: 영업, 3 : 폐업)
TRDSTATENM	영업상태명
DCBYMD	폐업일자
SITEAREA	소재지면적
RDNWHLADDR	도로명주소
BPLCNM	사업장명
UPTAENM	업태구분명
X	좌표정보(X)
Y	좌표정보(Y)

다. 주변 상권/ 교통시설 데이터 수집

카카오 api 서비스를 통해 각각의 데이터를 수집한다.

대학교에 대한 정보는 카페 주변 반경 1km 내 대학교 개수(num_unv), 가장 가까운 대학교 이름(unv_name)과 카페 사이의 직선 거리(unv_dist)를 불러와 cfdata에 병합하였다.

PVPERMYMD	TRDSTATEGBN	TRDSTATENM	DCBYMD	SITEAREA	RDNWHLADDR	BPLCNM	X	Y	region	num_unv	unv_name	unv_dist
2013-11-13	3	폐업	2017-10-11	22.53	서울특별시 강동구 양재대로87길 11 (성내동)	탐플레이커피	127.134306	37.525438	강동구	1	한국체육대학교	662
2015-07-08	3	폐업	2017-07-10	22.00	서울특별시 강동구 진왕도로27길 23, 1층 (천호동)	아카페	127.131877	37.539416	강동구	0	없음	x
2014-03-20	3	폐업	2017-07-17	74.21	서울특별시 강동구 구천면로 428 (명일동, 명일센트로빌 205호)	CAFE TENUTO 테누토	127.146158	37.550130	강동구	0	없음	x
2014-12-05	3	폐업	2017-06-02	337.19	서울특별시 강동구 강동대로 205, 1층 (성내동)	할리스커피 올림피아 공원북문점	127.129210	37.524394	강동구	1	한국체육대학교	490
서울특별시 강도												

공공시설에 대한 정보는 카페 주변 반경 1km 내 공공기관의 개수(num_po3)를 count 하여 병합하였다.

	APVPERMYMD	TRDSTATEGBN	TRDSTATENM	DCBYMD	SITEAREA	RDNWHLADDR	BPLCNM	X	Y	region	num_po3
0	2013-11-13	3	폐업	2017-10-11	22.53	서울특별시 강동구 양재대로87길 11 (성내동)	탭플레이 커피	127.134306	37.525438	강동구	8
1	2015-07-08	3	폐업	2017-07-10	22.00	서울특별시 강동구 진향도로27길 23, 1층 (천호동)	아카페	127.131877	37.539416	강동구	10
2	2014-03-20	3	폐업	2017-07-17	74.21	서울특별시 강동구 구천대로 428 (명일동, 명일센트로빌 205호)	CAFE TENUTO 테누토	127.146158	37.550130	강동구	11
3	2014-12-05	3	폐업	2017-06-02	337.19	서울특별시 강동구 강동대로 205, 1층 (성내동)	할리스커피 올림픽공원북문점	127.129210	37.524394	강동구	9
4	2013-01-03	3	폐업	2017-05-18	30.00	서울특별시 강동구 풍성로63길 84 (둔촌동)	도우드림	127.140339	37.531725	강동구	10

관광명소 데이터 또한 공공기관과 동일하게 개수만 count 하여 num_at4의 컬럼명으로 병합, 저장하였다.

	APVPERMYMD	TRDSTATEGBN	TRDSTATENM	DCBYMD	SITEAREA	BPLCNM	X	Y	region	ADDR	num_at4
0	2013-01-09	3	폐업	2017-08-31	48.84	일육커피	127.133821	37.545801	강동구	서울특별시 강동구 구천대로 292-1	6
1	2014-08-11	3	폐업	2017-05-08	3.30	블럭킹	127.140386	37.555312	강동구	서울특별시 강동구 고덕로 133	3
2	2014-04-03	3	폐업	2017-11-03	27.00	지희(ZIHEE)	127.131589	37.527617	강동구	서울특별시 강동구 양재대로89길 74	6
3	2014-04-03	3	폐업	2017-09-08	71.40	그리고카페(grigo cafe)	127.133850	37.537172	강동구	서울특별시 강동구 진향도로 66	8
4	2015-04-21	3	폐업	2017-10-19	42.90	구앤타오주스 강동성심점	127.135853	37.535925	강동구	서울특별시 강동구 천호대로173길 18	7
...

마지막으로 교통시설 정보이다. 카페 주변의 지하철역 기준으로 교통시설 데이터를 수집하였다. 카페 주변 반경 1km 내 지하철역 개수(sw8_unv), 가장 가까운 지하철역 명(sw8_name)과 카페 사이의 직선 거리(sw8_dist)를 불러와 cfdata에 병합하였다.

	APVPERMYMD	TRDSTATEGBN	TRDSTATENM	DCBYMD	SITEAREA	BPLCNM	X	Y	region	num_sw8	sw8_name	sw8_dist
0	2013-11-13	3	폐업	2017-10-11	22.53	탭플레이 커피	127.134306	37.525438	강동구	2	둔촌동역 5호선	298
1	2015-07-08	3	폐업	2017-07-10	22.00	아카페	127.131877	37.539416	강동구	4	강동역 5호선	394
2	2014-03-20	3	폐업	2017-07-17	74.21	CAFE TENUTO 테누토	127.146158	37.550130	강동구	3	명일역 5호선	234
3	2014-12-05	3	폐업	2017-06-02	337.19	할리스커피 올림픽공원북문점	127.129210	37.524394	강동구	3	둔촌동역 5호선	717
4	2013-01-03	3	폐업	2017-05-18	30.00	도우드림	127.140339	37.531725	강동구	4	둔촌동역 5호선	580

라. 공시지가 데이터 수집

서울 열린 데이터 광장의 '서울시 개별공시지가 정보'로 API로 제공받아 2020년 기준의 공시지가 정보(JIGA)를 수집하였다.

	APVPERMYMD	TRDSTATEGBN	TRDSTATENM	DCBYMD	SITEAREA	BPLCNM	X	Y	region	ADDR	JIGA
0	2013-11-13	3	폐업	2017-10-11	22.53	탐플레이커피	127.134306	37.525438	강동구	서울특별시 강동구 양재대로 87길 11	6350000
1	2015-07-08	3	폐업	2017-07-10	22.0	아카페	127.131877	37.539416	강동구	서울특별시 강동구 진왕도로 27길 23	4831000
2	2014-03-20	3	폐업	2017-07-17	74.21	CAFE TENUTO 테누토	127.146158	37.550130	강동구	서울특별시 강동구 구천면로 428	6596000
3	2014-12-05	3	폐업	2017-06-02	337.19	할리스커피 올림픽 공원북문	127.129210	37.524394	강동구	서울특별시 강동구 강동	8360000

마. 유동인구 데이터 수집

서울 열린 데이터 광장에의 '서울특별시 주민등록인구 통계' 데이터이다. 2017년 1분기부터 2021년 2분기까지의 데이터로 2021년 3분기 이후의 카페 데이터는 2021년 2분기 데이터로 적용하였다.

cfdata.iloc[:,17:]

	계	0~4세	5~9세	10~14세	15~19세	20~24세	25~29세	30~34세	35~39세	40~44세	...	55~59세	60~64세	65~69세	70~74세	75~79세	80~84세	85~89세	90~94세	95~99세	100세 이상
0	24517	924	786	916	1289	1709	1883	1980	2252	1877	...	2081	1630	1136	852	560	276	129	46	14	10
1	25687	869	624	633	949	1540	2079	2417	2261	1825	...	2528	2090	1426	1052	694	374	162	74	17	13
2	22718	822	929	1195	1588	1573	1474	1496	1695	1803	...	1929	1593	1045	674	412	282	156	80	13	7
3	24336	945	809	887	1232	1718	1863	2029	2296	1889	...	2041	1603	1126	820	543	266	122	42	15	9
4	28628	1172	1140	1182	1655	2024	2099	2162	2400	2001	...	2613	2085	1325	838	552	349	201	72	15	15

바. 경쟁업체 데이터 전처리

두 가게의 거리가 1km미만일 때 개·폐업 여부와 폐업시기 등을 비교할 수 있는 데이터를 만들어 본래 cfdata에 추가하였다.

DCBYMD	SITEAREA	BPLCNM	X	Y	region	ADDR	...	op_cl_6	op_cl_3	cl_op	cl_op_12	cl_op_6	cl_op_3	cl_cl	cl_cl_12	cl_cl_6	cl_cl_3
2017-10-19	25.10	커피나무	126.920799	37.497518	동작구	서울특별시 동작구 여의대방로22나길 1	...	0	0	40	10	5	0	65	0	0	0
2017-05-16	154.62	카페아이애틀송실대점	126.954995	37.495140	동작구	서울특별시 동작구 상도로 381	...	0	0	24	5	2	1	45	0	0	0
2017-07-19	20.00	그리다꿈노랑진점	126.944998	37.512912	동작구	서울특별시 동작구 만양로14가길 30-3	...	0	0	63	11	6	1	114	0	0	0
						서울...											

사. 확진자 데이터 전처리

일 별로 기록되어 있는 데이터의 매 달 마지막 날의 누적 확진자 수에서 전 달까지의 확진자 수를 뺀 월 별 확진자 데이터 프레임으로 정리한다. 서울시 전체 데이터는 구 별 누적 확진자 수를 모두 더한 후 위의 방법을 사용하였다.

Year	Month	confirmed	confirmedadd
2020	2	82	82
2020	3	372	454
2020	4	179	633
2020	5	228	861
2020	6	452	1313
2020	7	288	1601
2020	8	2314	3915
2020	9	1379	5294
2020	10	718	6012
2020	11	2795	8807
2020	12	10222	19029
2021	1	5133	24162
2021	2	4068	28230
2021	3	3801	32031
2021	4	5806	37837
2021	5	6080	43917
2021	6	6070	49987
2021	7	14372	64359
2021	8	14995	79354
2021	9	21141	100495

서울시 확진자(confirmed:누적, confirmedadd:추가)

Year	Month	confirmed	confirmedadd
2020	2	1	1
2020	3	22	23
2020	4	8	31
2020	5	9	40
2020	6	13	53
2020	7	17	70
2020	8	105	175
2020	9	59	234
2020	10	28	262
2020	11	124	386
2020	12	406	792
2021	1	215	1007
2021	2	150	1157
2021	3	172	1329
2021	4	275	1604
2021	5	220	1824
2021	6	220	2044
2021	7	726	2770
2021	8	674	3444
2021	9	705	4149

동작구 확진자(confirmed:누적, confirmedadd:추가)

아. 서울시 사회적 거리두기 단계별 정보 수집

- 2020년 2월 20일 신천지 확진자 발생
- 2020년 2월 29일 사회적 거리두기 제안
- 2020년 3월 22일 ~ 강화된 사회적 거리두기(1단계)
- 2020년 8월 19일 ~ 거리두기 2단계
- 2020년 10월 12일 ~ 거리두기 1단계
- 2020년 11월 19일 ~ 거리두기 1.5단계
- 2020년 12월 1일 ~ 12월 7일 거리두기 2단계 (카페 착석 금지)
- 2020년 12월 8일 ~ 2021년 2월 14일 거리두기 2.5단계
(사적 모임 4인까지, 카페 착석 금지, 관광명소 폐쇄)
- 2021년 2월 15일 ~ 7월 12일 2단계
(사적 모임 4인까지, 카페 22시 이후 포장배달)
- 2021년 7월 12일 ~ 4단계

IV 데이터 분석

1. 분석용 데이터 준비

가. 수집 및 전처리 데이터 병합

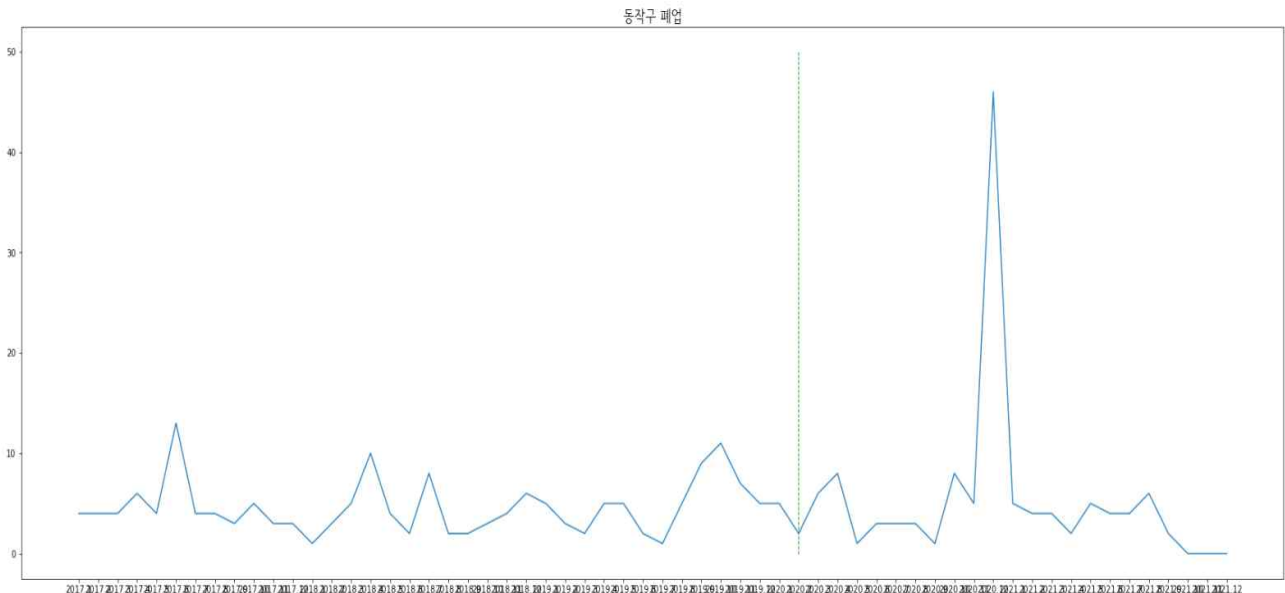
앞서 수집 및 전처리를 통해 준비한 데이터들을 모두 병합해준 후 데이터 정보를 확인합니다.

```
cfdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26545 entries, 0 to 26544
Data columns (total 58 columns):
#   Column                Non-Null Count  Dtype
---  -
0   APVPERMYMD            26545 non-null  object
1   TRDSTATEGBN          26545 non-null  int64
2   TRDSTATENM           26545 non-null  object
3   DCBYMD               10145 non-null  object
4   SITEAREA             25600 non-null  object
5   BPLCNM              26545 non-null  object
6   X                   26545 non-null  float64
7   Y                   26545 non-null  float64
8   region              26545 non-null  object
9   ADDR                26545 non-null  object
10  JIGA                26545 non-null  int64
11  op_op              26545 non-null  int64
12  op_op_12           26545 non-null  int64
13  op_op_6            26545 non-null  int64
14  op_op_3            26545 non-null  int64
15  op_cl              26545 non-null  int64
16  op_cl_12           26545 non-null  int64
17  op_cl_6            26545 non-null  int64
18  op_cl_3            26545 non-null  int64
19  cl_op              26545 non-null  int64
20  cl_op_12           26545 non-null  int64
21  cl_op_6            26545 non-null  int64
22  cl_op_3            26545 non-null  int64
23  cl_cl              26545 non-null  int64
24  cl_cl_12           26545 non-null  int64
25  cl_cl_6            26545 non-null  int64
26  cl_cl_3            26545 non-null  int64
27  region_2           26545 non-null  object
28  계                  26545 non-null  object
29  0~4세              26545 non-null  object
30  5~9세              26545 non-null  object
31  10~14세            26545 non-null  object
32  15~19세            26545 non-null  object
33  20~24세            26545 non-null  object
34  25~29세            26545 non-null  object
35  30~34세            26545 non-null  object
36  35~39세            26545 non-null  object
37  40~44세            26545 non-null  object
38  45~49세            26545 non-null  object
39  50~54세            26545 non-null  object
40  55~59세            26545 non-null  object
41  60~64세            26545 non-null  object
42  65~69세            26545 non-null  object
43  70~74세            26545 non-null  object
44  75~79세            26545 non-null  object
45  80~84세            26545 non-null  object
46  85~89세            26545 non-null  int64
47  90~94세            26545 non-null  int64
48  95~99세            26545 non-null  int64
49  100세 이상         26525 non-null  object
50  num_sw8            26545 non-null  int64
51  sw8_name           26545 non-null  object
52  sw8_dist           26545 non-null  object
53  at4                26545 non-null  int64
54  po3                26545 non-null  int64
55  num_unv            26545 non-null  int64
56  unv_name           26545 non-null  object
57  unv_dist           26545 non-null  object
dtypes: float64(2), int64(25), object(31)
memory usage: 11.7+ MB
```

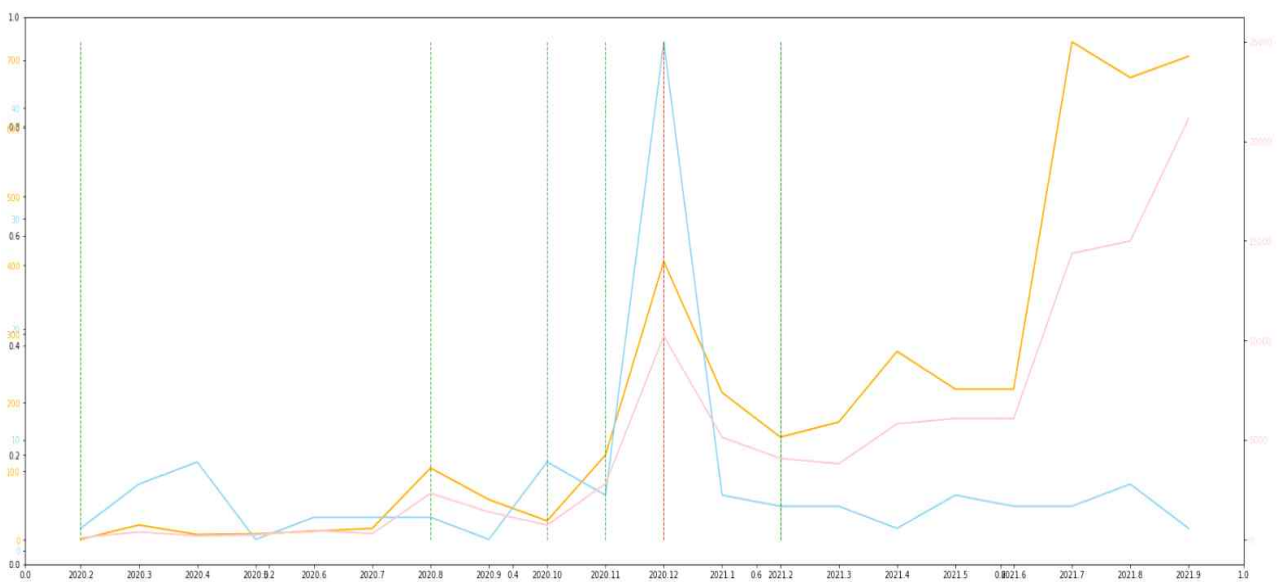
2. 탐색적 분석

가. 동작구 폐업 가게 수 시각화



2020년 2월 첫 사회적 거리두기 권고(초록색 점선) 이후 큰 변화 없었으나 2020년 12월 폐업 가게 수가 급격히 증가함을 볼 수 있다.

이를 자세히 보기 위하여 서울시 확진자 수와 동작구 확진자 수를 포함하여 사회적 거리두기 단계에 대한 정보를 포함한 그래프를 그려보았다.

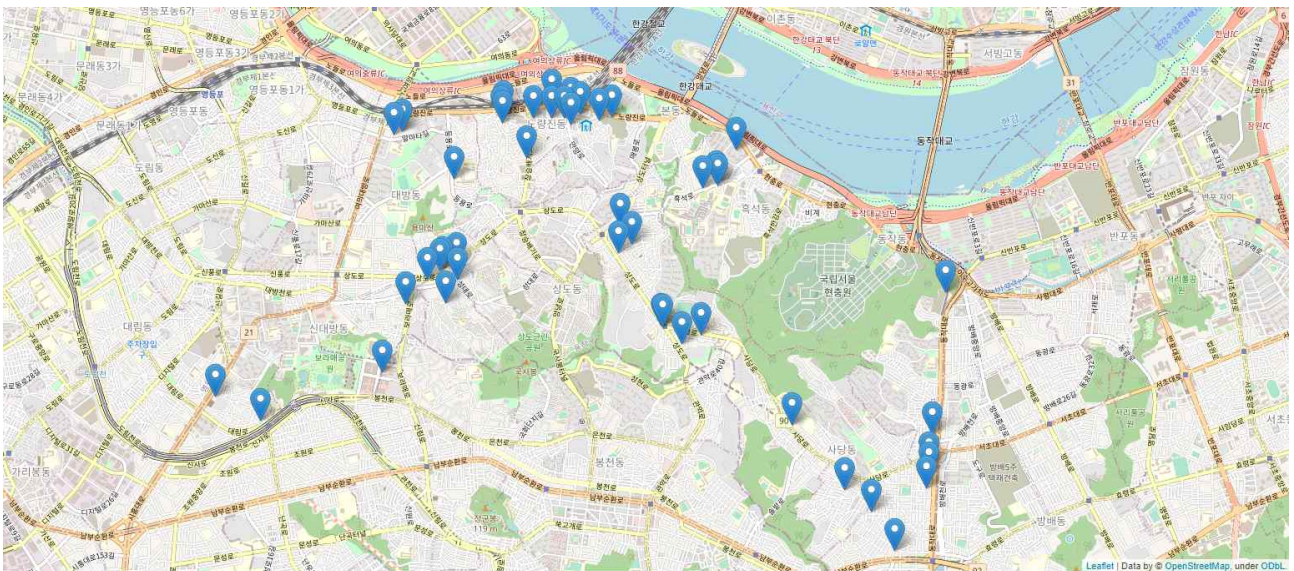


(분홍색 그래프:서울시 확진자 수, 주황색 그래프:동작구 확진자 수, 주황색 그래프:동작구 폐업가게 수, 초록색 세로 점선: 사회적 거리두기 단계 변화, 빨간색 세로 점선:사회적 거리두기 2단계, 2.5단계)

위 그래프를 확인해보면 2020년 12월 확진자 수가 급증하고, 폐업한 가게 수 역시 급증하였다. 하지만 2021년 이후에는 확진자 수가 꾸준히 증가함에도 폐업하는 가게가 크게 늘지 않았다. 이는 2020년 12월 시행된 사회적 거리두기 2단계와 2.5단계 조항에 '커피전문점 내 착석 금지' 조항의 영향이 컸을 것으로 예상된다.

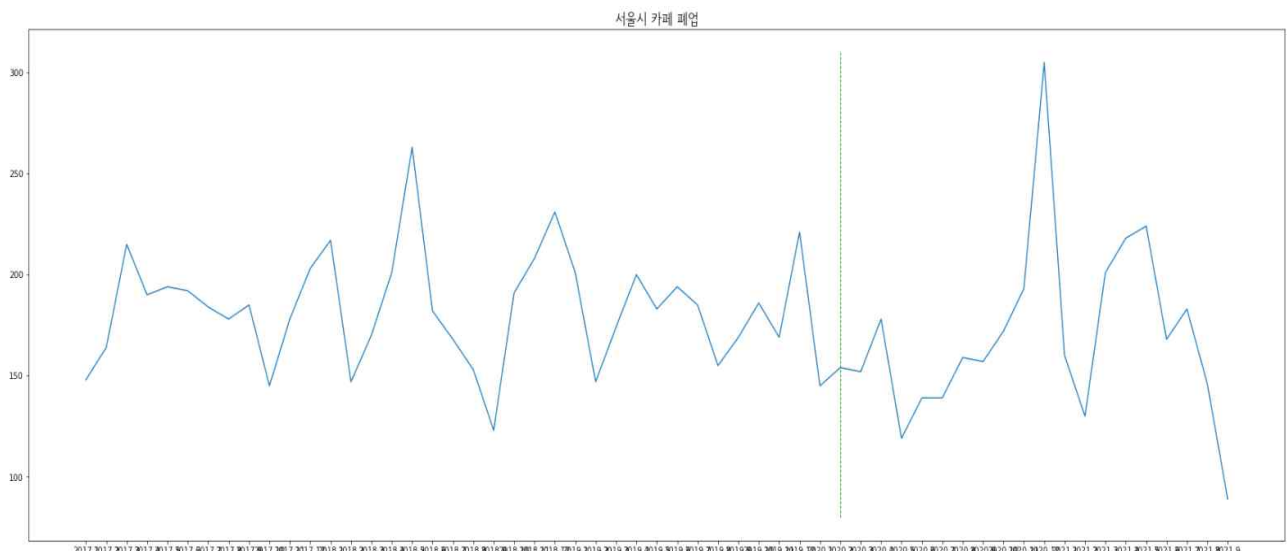
나. 동작구 카페 map 시각화

위의 그래프에서 확진자 수와 폐업 가게 수가 급격하게 늘었던 2020년 12월 폐업한 가게의 지도 시각화를 하여 카페의 위치적 특성을 확인해 본다.

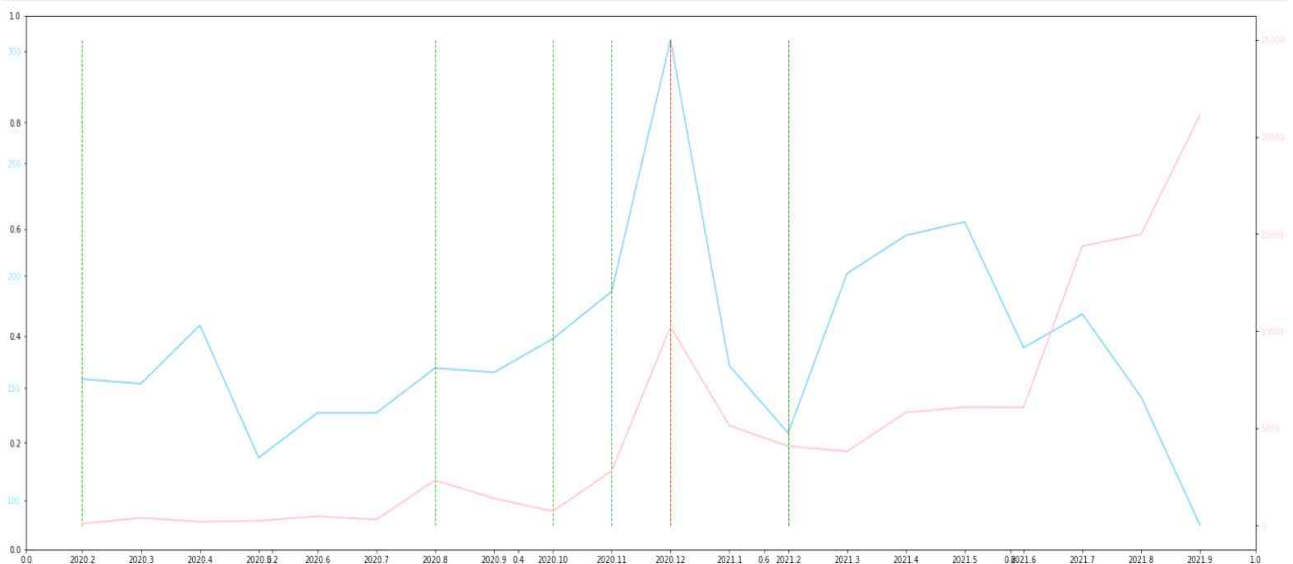


노량진역, 신대방역, 이수역, 송실대학교, 중앙대학교 주변의 카페들이 많이 폐업한 것을 알 수 있다.

다. 서울특별시 전체 폐업 카페 시각화

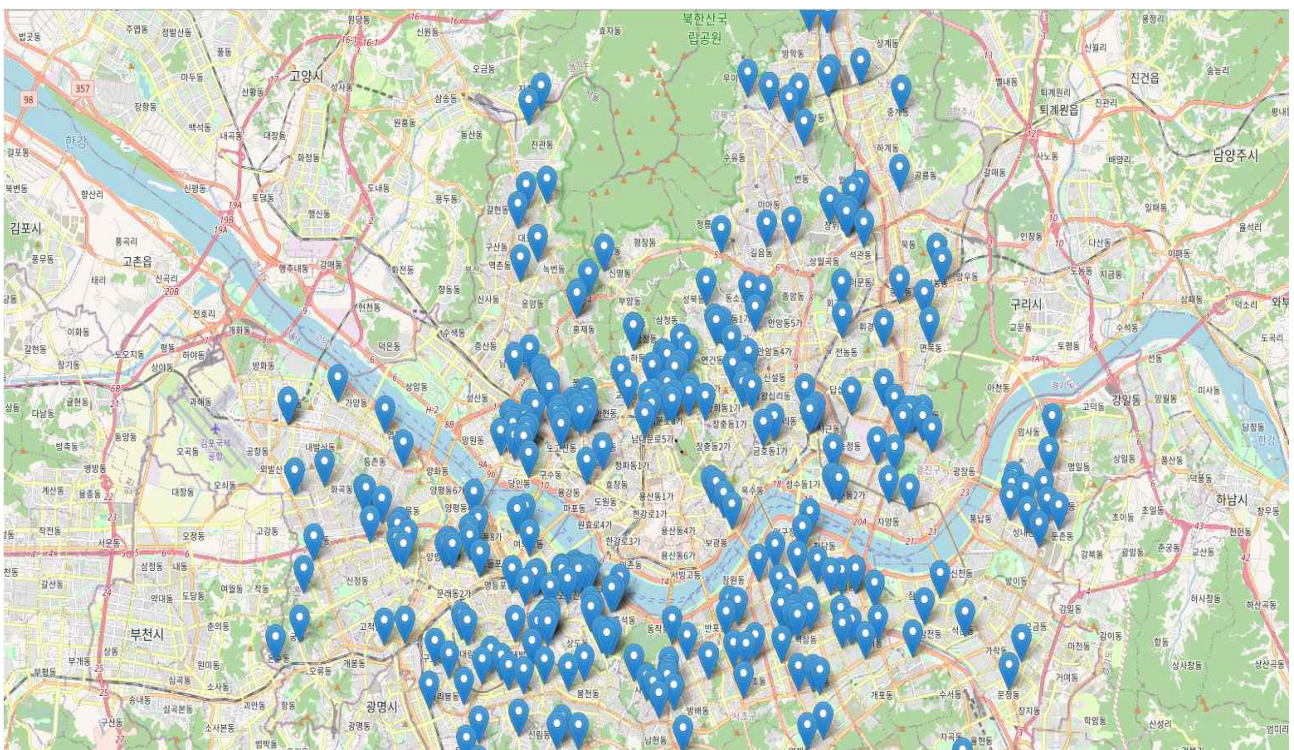


(2017년1월~2021년9월 서울시 내 폐업 카페 수)



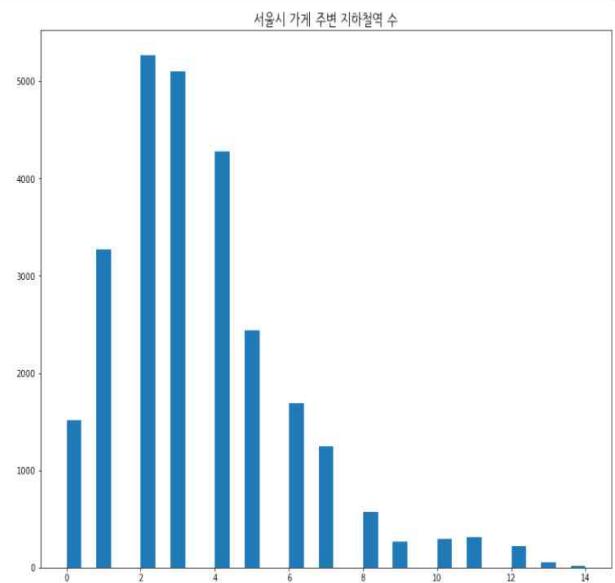
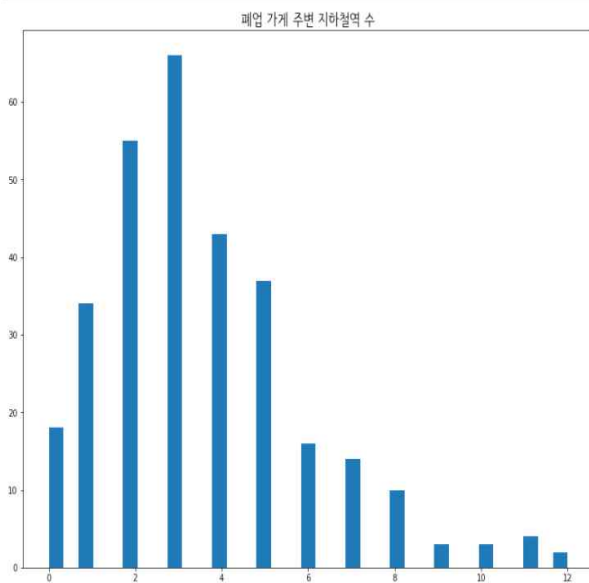
(파란색 그래프:서울시 폐업 가게 수, 분홍색 그래프:서울시 확진자 수)

서울 시 전체 그래프를 확인해보면 확진자 수는 카페 폐업에 크게 영향을 미치지 않은 것을 더 확실히 알 수 있다.

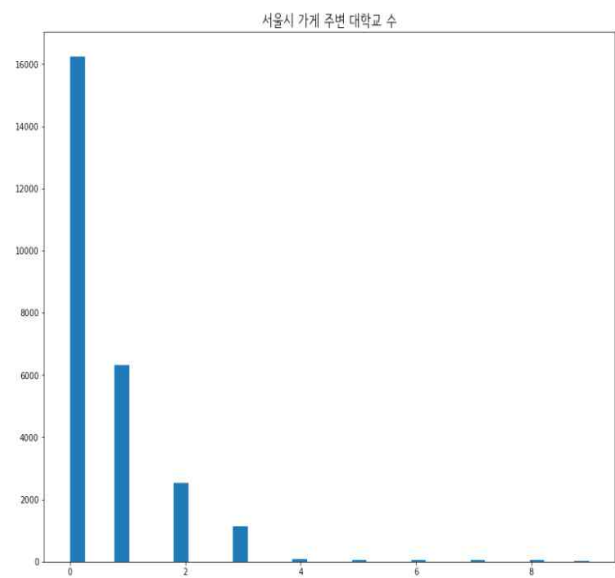
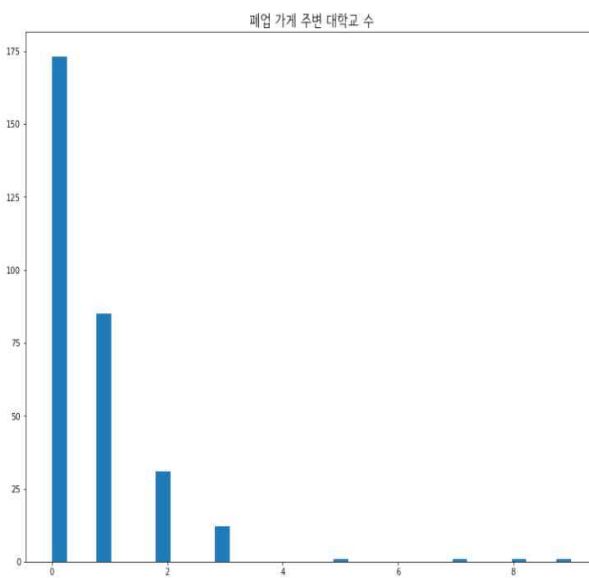


2020년 12월 서울특별시 내 폐업한 카페의 지도 시각화이다. 이들은 대부분 지하철역 주변, 대학교 주변에 분포한 것을 확인할 수 있다.

라. 가게 주변 정보의 히스토그램



카페 주변의 지하철역의 수가 3개일 때 다른 가게들에 비해 폐업한 가게가 많음을 확인할 수 있다.



카페 주변의 대학교의 수가 1개일 때 다른 가게들에 비해 폐업한 가게가 많음을 확인할 수 있다.

3. 모델링

가. 모델링 전 데이터 전처리

- ☐ 분석을 위하여 데이터의 누락값과 타입을 확인한 후 전처리한다.
- ☐ 문자열과 float형 타입이 섞여있는 'SITEAREA' 컬럼을 우선 모두 문자열로 변환 후 반복문과 locale을 사용하여 float형으로 변환한다.
- ☐ 데이터 세트를 훈련 데이터 세트와 테스트 데이터 세트로 나눈다.(7:3)

```
cf1 = cfdata.iloc[:, [1,4]]
cf2 = cfdata.iloc[:, 10:27]
cf3 = cfdata.iloc[:, [50,53,54,55]]
cf = pd.concat([cf1,cf2,cf3],axis=1)
cf = cf.reset_index(drop=True)
cf.columns

Index(['TRDSTATEGBN', 'SITEAREA', 'JIGA', 'op_op', 'op_op_12', 'op_op_6',
      'op_op_3', 'op_cl', 'op_cl_12', 'op_cl_6', 'op_cl_3', 'cl_op',
      'cl_op_12', 'cl_op_6', 'cl_op_3', 'cl_cl', 'cl_cl_12', 'cl_cl_6',
      'cl_cl_3', 'num_sw8', 'at4', 'po3', 'num_unv'],
      dtype='object')

cf = cf.dropna()

from sklearn.model_selection import train_test_split

X, y = cf.iloc[:, 1:].values, cf.iloc[:, 0].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0, stratify=y)
```

나. 다중선형회귀모델링

- ☐ LinearRegression을 사용하여 모델링을 진행한다.

```
from sklearn.linear_model import LinearRegression

mlr = LinearRegression()

mlr.fit(X_train, y_train)
```

4. 모델링 평가

가. 예측 정확도 확인

```
y_train_pred = mlr.predict(X_train)
y_test_pred = mlr.predict(X_test)

from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error

print('훈련 MSE: %.3f, 테스트 MSE: %.3f' % (
    mean_squared_error(y_train, y_train_pred),
    mean_squared_error(y_test, y_test_pred)))
print('훈련 R^2: %.3f, 테스트 R^2: %.3f' % (
    r2_score(y_train, y_train_pred),
    r2_score(y_test, y_test_pred)))
```

훈련 MSE: 0.312, 테스트 MSE: 0.314

훈련 R^2: 0.670, 테스트 R^2: 0.669

□ 훈련 세트와 테스트 세트의 예측에 대한 MSE(평균 제곱 오차)가 각각 0.312, 0.314로 모델이 과대적합 되지 않고 적절히 모델링되었음을 알 수 있습니다.

□ 훈련 세트와 테스트 세트의 결정계수는 각각 0.670, 0.669로 모델이 아주 잘 데이터를 설명하지는 못하나 어느정도 설명함을 알 수 있습니다.

나. 결과 해석

```
print('x 변수: ', cf.columns[1:])
print('\n')
print('x 변수의 기울기: ', mlr.coef_)
print('\n')
print('절편: ', mlr.intercept_)

x 변수: Index(['SITEAREA', 'JIGA', 'op_op', 'op_op_12', 'op_op_6', 'op_op_3', 'op_cl',
              'op_cl_12', 'op_cl_6', 'op_cl_3', 'cl_op', 'cl_op_12', 'cl_op_6',
              'cl_op_3', 'cl_cl', 'cl_cl_12', 'cl_cl_6', 'cl_cl_3', 'num_sw8', 'at4',
              'po3', 'num_unv'],
              dtype='object')

x 변수의 기울기: [-1.45562051e-06  1.46342230e-10 -2.63444768e-03 -3.83800523e-03
 1.13882174e-03 -4.23362311e-03  1.19134799e-03 -1.11478863e-03
-1.56552229e-03  2.48667221e-03 -3.01164546e-03  6.41302042e-03
-3.18993693e-03  6.87798444e-03  5.79262998e-03 -1.30104261e-18
 0.00000000e+00 -4.33680869e-19  5.32130482e-04  2.39443734e-03
 2.94467628e-03 -9.06407862e-03]

절편: 1.7125721934273033
```

□ 데이터의 열을 살펴보면 대학교의 수(num_unv), 폐업한 가게 주변 3개월 내 영업 중인 가게 수(cl_op_3), 폐업한 가게 주변 12개월 내 영업 중인 가게 수(cl_cl_12)가 가게의 영업/폐업에 영향을 주었음을 알 수 있다.

□ 탐색적 데이터 분석에서 예상하였던 카페 주변 대학교의 수가 큰 영향이 있었음을 확인할 수 있었으며 앞서 예상하지 못하였던 폐업한 가게 주변의 경쟁업체에 대한 영향을 알 수 있었다.

V 평가 및 전개

1. 프로젝트 평가 및 향후 계획

□ 탐색적 분석을 진행하며 생각보다 카페의 폐업에 사회적 거리두기의 영향이 많이 컸음을 알아 이 주제의 분석 프로젝트가 좋은 결과가 나올지에 대해 걱정이 있었고 모델링 후 결과도 데이터 수집량에 비해 정확도가 떨어진 점이 아쉽다.

□ 데이터 수집과 전처리에 많은 공을 들였으나 이로 인하여 분석에 대한 시간이 많이 부족했던 점이 아쉽다. 이후 다른 프로젝트를 진행하게 된다면 시간 분배에 대한 계획을 체계적으로 작성할 예정이다.

참 고 문 헌

- 금융빅데이터플랫폼(비씨카드), "거리두기 전후의 소비변화",2021.02.06.
- 금융빅데이터플랫폼(비씨카드), "코로나 전후의 소비변화",2021.02.16.
- 정동규,윤희연, "발달상권과 골목상권에 위치한 음식점의 생존과 폐업 비교", 대학건축학회논문집-계획계, Vol.33 No.3, 2017, p.57-68.
- 이충수, "외식창업 성공요인 분석", 예술인문사회융합멀티미디어논문지, Vol.8 No.11, 2018, p.171-180.
- 방준아,손광민,이소정,이현근,조수빈, "서울 치킨집 폐업 예측 모형 개발 연구", 한국빅데이터학회 학회지,Vol.3 No.2, 2018, p.35-49
- 유현지 , "코로나19와 서울시 골목상권의 매출액 영향요인에 관한 연구", 한국지역개발학회지, 33(3), 2021, 45-75
- 주재욱, 노승철, 윤종진 , "코로나19 확산이 서울 지역에 미친 경제적 손실", 정책리포트, 2021, 1-22
- "3달째 이어지는 '5인 이상 집합금지'..."버티지못해 폐업"
 , <https://www.fnnews.com/news/202103121216394942>
- "영업재개 했지만, 매출은 9천원... 카페, 헬스장의 '눈물'"
 ,
http://www.ohmynews.com/NWS_Web/View/at_pg.aspx?CNTN_CD=A0002712985&CMP_T_CD=P0010&utm_source=naver&utm_medium=newsearch&utm_campaign=naver_news

