



# 편의점 **최적입지** 선정을 위한 조건

서울시 편의점 생존율 예측을 통한 상권 분석



## 프로젝트 소개

분석 배경  
분석 목적  
데이터 설명



## 분석 과정

데이터 시각화  
데이터 전처리  
회귀분석



## 분석 결과

머신러닝



## 결론

생존분석  
입지추천  
시사점 및 한계점



# 프로젝트 소개

- 분석 배경
- 분석 목적
- 데이터 설명

로이칼리언이론 뉴스시스  
**NEWSIS**

전국 편의점, 지난해 4만3632개...100m 내 신규 점포 늘어

기사입력 2020.10.05. 오후 4:08

기사원문 스크랩

문들기 · 설정

[표] 지역별 편의점 간 거리평균(m)

[표] 서울지역 편의점 간 거리평균(m)

지역	최단거리(m)
전체 평균	224.9
서울	104.6
부산	148.8
대전	150.4
광주	157.3
대구	168.5
울산	194.3
인천	202.2
경기	212.9
세종	223.2
제주	266.9
경남	288.4
충북	322.4
경북	327.6
충남	341.2
전북	343.6
강원	367.6
전남	488.3

지역	최단거리(m)
전체 평균	104.6m
충구	75.8
충로구	87.4
강남구	92.7
마포구	94.3
관악구	96.3
동대문구	96.6
동작구	98.5
금천구	102.8
광진구	104
성동구	104.6
강서구	106.4
영등포구	106.7
서대문구	107.6
강북구	108.2
강동구	108.8
송파구	111.1
구로구	112.9
서초구	113.3
성북구	115.1
중랑구	115.4
용산구	115.6
노원구	116.9
양천구	117.5
도봉구	120.6
은평구	122

MT  머니투데이

"알바생보다 수입 적다"는데...편의점 왜 계속 늘까

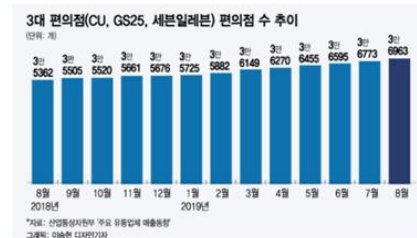
기사입력 2019.10.24 오후 12:10 | 최종수정 2019.10.24 오후 12:41 | 기사원본 |  SNS |  로그인 |  로그아웃

4 4

요약본 가  

[머니투데이 최성근 이코노미스트] [편집자주] 복잡한 경제 이슈에 대해 단순한 해법을 모색해 봅니다.

[[소프트 랜딩]편의점 증가하면 점포당 수익성 악화하지만 몰사 매출은 증가]



최근 '저렴함-저성능'으로 디플레이션 우려까지 나올 정도로 경기는 불황이지만에도 오히려 편의점은 늘고 있다. 산업통상자원부에 따르면 2019년 8월 기준 국내 3대 편의점(CU, GS25, 세븐일레븐) 점포수는 총 3만9666개로 지난해 8월 3만9362개보다 16000개(4.5%) 증가한 것으로 나타났다. 여기에 신생 편의점인 이마트24(4290개)와 일련제 1500개에 소유했던 미니스톱(2574개)까지 포함하면 현재 편의점 점포수는 대략 4만4800개에 육박할 것으로 추정된다.

한국 편의점산업협회에 따르면 2018년 기준으로 편의점 1개당 인구수는 1226명으로 '편의점 왕국'이라 불리는 일본의 2249명과 비교해 2배 가량 높은 밀집도를 나타내고 있다. 이처럼 세계에서 가장 편의점 밀집도가 높음에도 오히려만 3대 편의점들 기준으로 1월~8월까지 12380여 증가한 점도를 점포수는 계속해서 증가하고 있다.

더욱이 올해는 최저임금이 10.9% 인상되면서 패이전 전조들이 하루 종일 일해도 아

머니투데이 랭킹 뉴스

오후 2시~오후 3시까지 집계한 결과입니다.

1. '엄마가 미안해' 울음바다 된 '보  
습 온전' 6세 사망 재 ...  
◎ 2시간전

2 여가부장관 '성인지 학술기원' 준  
환 받 "아이 키우기 좋아"

© 1998

선·트럼프 소동이 변수  
◎ 4시간전

4 또 급증한 김포 아파토값...서울  
전세 표기한 직장인이 사라졌...

5. 전세보증금 반환이 어렵게 되었다

오재일과함께  NAVER TV

스웨덴에서는 어른들의 오디오

**동시 생중계**

[illegible]

호의 흥남기 자신임에도, 그게 가뭇하는 세종판가

이낙연 "경제 일원화할 '후학'지망개발부' 신설 검토"중...

[이 대안] 4명만 남을 바이론...경우의 수 하나뿐인 트릴

전체난 문선교과여 전국 집값 올락...4개월만에 최고...

한국언어리본인 80년대, 1. 다문화 리본 3년 연속

카카오 3분기 1조 매출 1천억 영업이익 올라..연착륙 질..

교육정책의 변화에 대해 부정적 견해에 우세함

## 늘어만 가는 편의점

# I. 분석 배경

시사저널

✓ PICK ①

## 3분 거리에 편의점 7곳... '상생'은 없고 '상처'만 남았다

기사입력 2020.10.13. 오전 9:02 | 기사원문 | 스크린 | 본문듣기 · 설정

편의점 본사 매출은 느는데...

(단위:억원)

가맹점 평균매출은

ChosunBiz

2017년 2018년 2019년



## "편의점 과밀 출점 멈춰라"... 뿔난 편의점 가맹점주들

기사입력 2020.08.12. 오후 2:48 | 기사원문 | 스크린 | 본문듣기 · 설정

"담배 판매권 거리를 기준으로 매장 출점 허용해야"

신규 출점 제한, 기존 편의점 영업권만 보호하는 것 아니냐는 지적도 나와

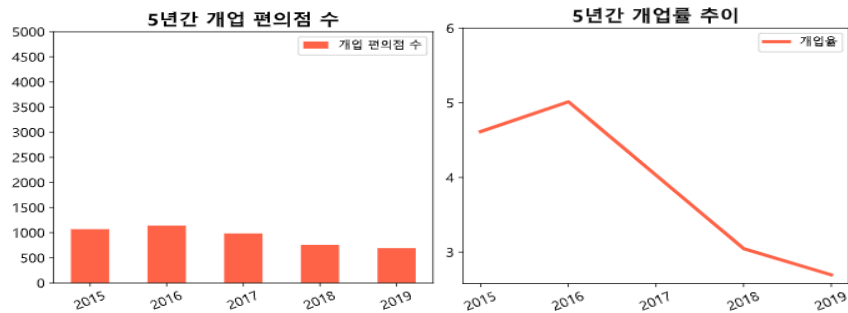


매출 저조로 인한 폐업점 증가

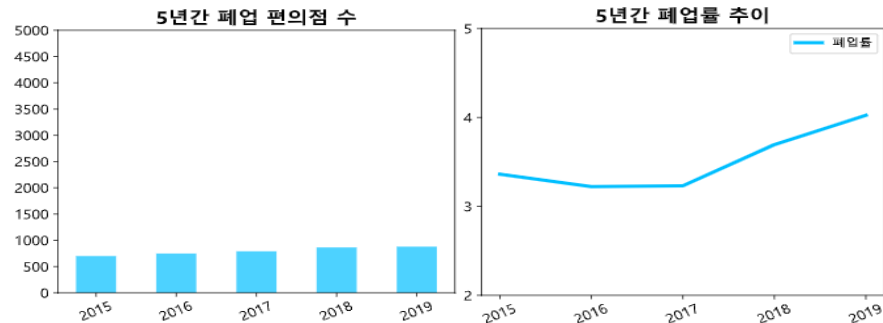


편의점 이미 포화상태

# I. 분석 배경



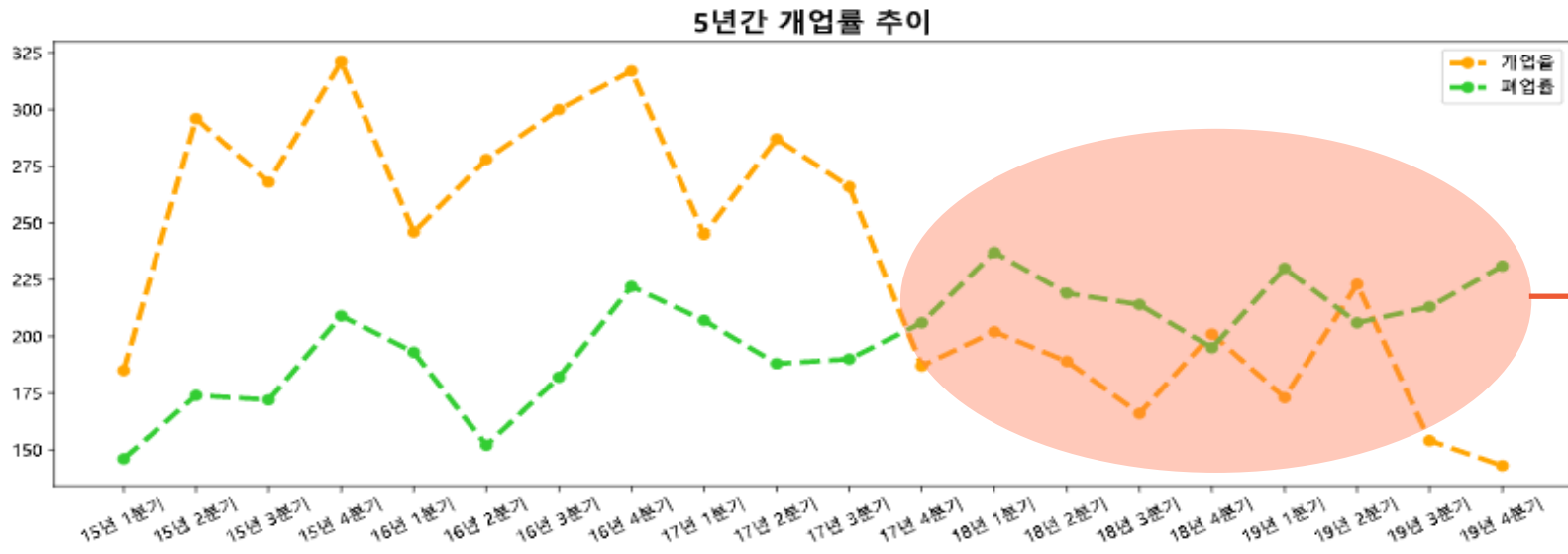
개업률 하락



폐업률 상승

5년간 서울시 공공데이터 분석 결과  
편의점의 **폐업률 역전** 양상을 보임

# I. 분석 목적



분기 별로 세분화하면 최근에 더욱 두드러지는 양상

편의점 창업을 위한 최적의 위치를 분석해  
입지 추천과, 생존율을 제시하자 !!

## I. 개발도구 소개

### 분석도구



NumPy



### 협업도구



NOTION  
아이디어 및 진행현황공유



Git hub  
코드관리 및 버전관리



# I . 사용 데이터

## 편의점매출

당월 매출 금액	분기별 도로명별 월 매출 (월단위)
요일별 매출 금액	분기별 (월단위)
요일별 매출 금액	분기별 (월단위)
연령대별 매출 금액	분기별 (월단위)
점포수	도로명별 점포 수

## 편의점 개·폐업 정보

일반 편의점 점포 수	분기별 도로명별 프랜차이즈 점포 수
프랜차이즈 점포 수	분기별 도로명별 총 편의점 점포 수
총 편의점 점포 수	분기별 도로명별 총 편의점 점포 수
총 간이 술집 점포 수	분기별 도로명별 총 술집 점포 수

출처 : 서울시 열린데이터 광장

# I. 사용 데이터

## 집객시설 정보

집객시설 수	분기·도로명 별 집객 시설 총 수
각각의 집객시설	'관공서수', '은행수', '약국수' 등 20개

## 아파트

아파트 평균시가	분기·도로명별 평균 시가
면적별 세대 수	분기·도로명·면적별 세대 수
가격대별 세대 수	분기·도로명·가격대별 평균 시가

## 유동인구 및 상주인구

총 유동인구	분기·도로명별 총 유동인구
연령대별 유동인구	분기·도로명·연령대별 유동인구
요일대별 유동인구	분기·도로명·요일대별 유동인구
성별 간 유동인구	분기·도로명·성별 간 유동인구

출처 : 서울시 열린데이터 광장

# I. 사용 데이터

## 직장인구

총 직장인구	분기·도로명·연령대별직장인구
연령대별 유동인구	분기·도로명·요일대별직장인구
성별 간 유동인구	분기·도로명·성별간직장인구

## 파생 변수

생존률	$\frac{\text{총 점포수}}{\text{총 점포수} + \text{폐업점포수}} \times 100$
폐업률	$\left(1 - \frac{\text{총 점포수}}{\text{총 점포수} + \text{폐업점포수}}\right) \times 100$
평균 당월 매출 금액	$\frac{\text{도로명별 당월 매출}}{\text{도로명별 점포 수}}$
집객시설 수	의료기관, 교육기관, 대중교통

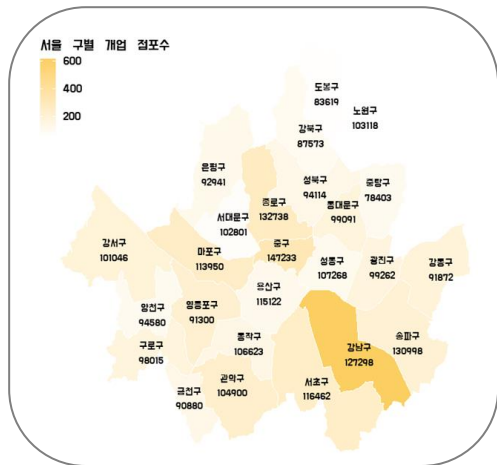
## 그 외 데이터

구별 임대료	구별 임대료
출처	우리마을가게상권분석 Crawling
GIS 포맷	2020년 05월 시군구별 행정구역 (SHP)
출처	딥러닝 기반 기술연구소 @지오서비스

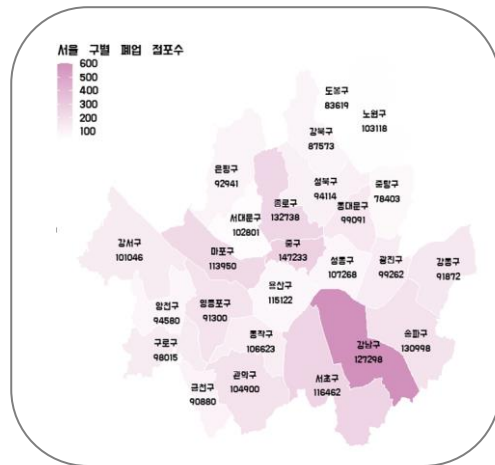
출처: 서울시 열린데이터 광장

## 1. 분석 목적

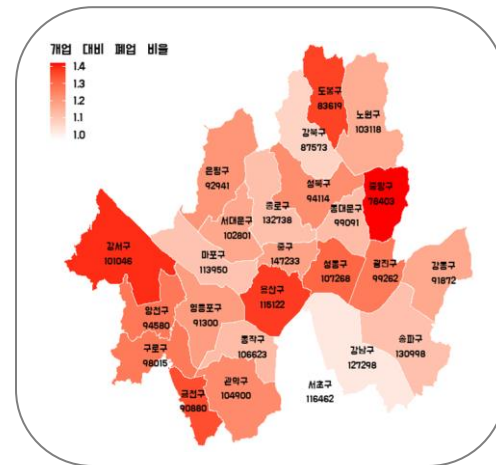
## 개업한 점포 수



## 폐업한 점포 수



개업수 대비 폐업수



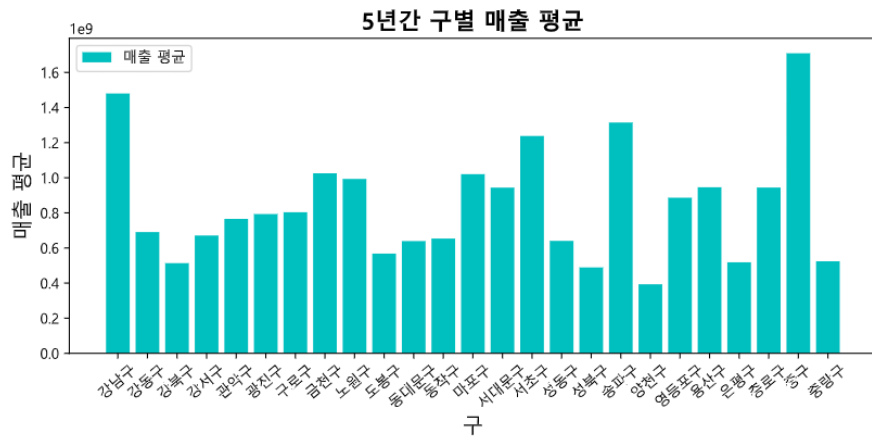
## R 과 대한민국 행정구역[SHP] 파일 활용 지도 시각화



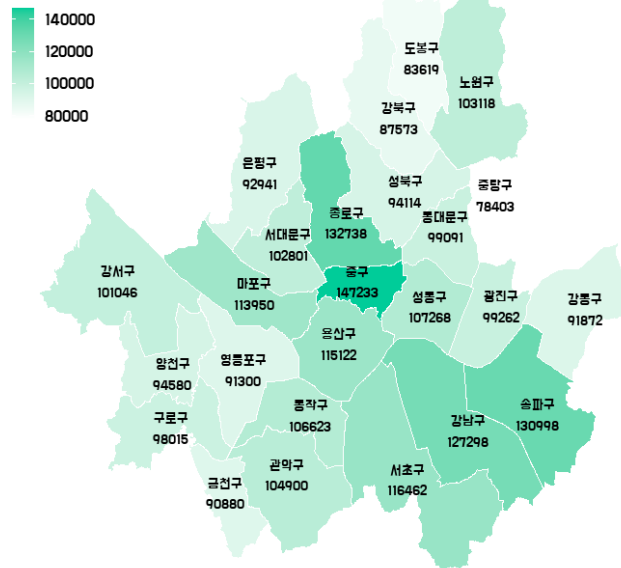
## 분석 과정

- 데이터 시각화
- 데이터 전처리
- 회귀분석

## II. 데이터 시각화

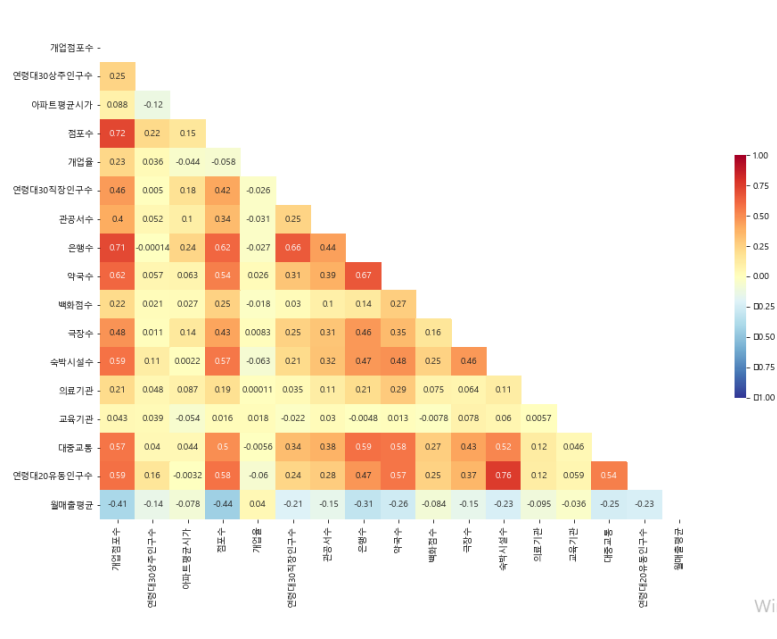
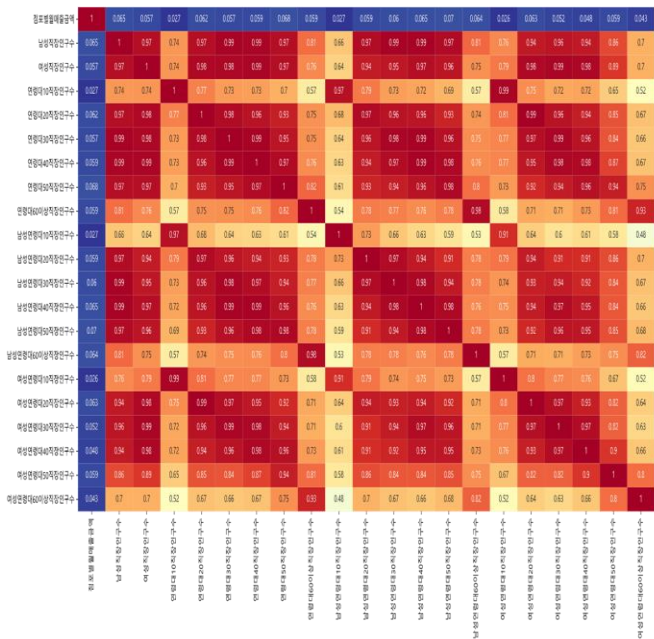


서울 구별 평균 임대료



각 시군구마다 상이한 **평균 매출액과 임대료**

## II. 데이터 전처리



히트맵을 통해 각 독립변수 간의 관계 파악  
비슷한 다수의 인구데이터로 높은 **다중공선성 발생**

종속변수(생존율)과 상관관계를 고려  
**전진 및 후진 제거법**으로 변수 선별

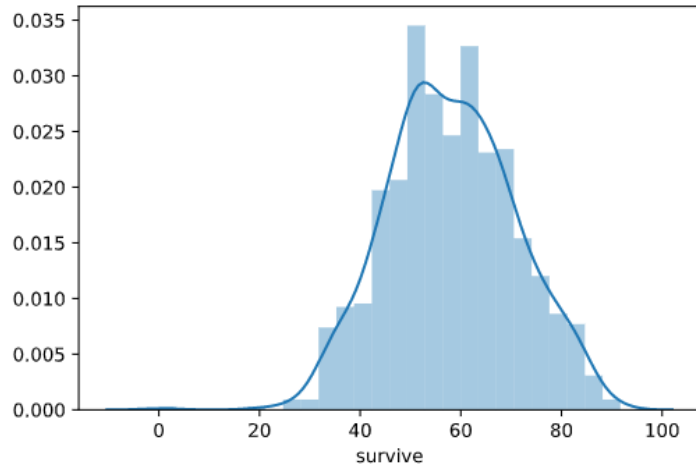
## II. 데이터 전처리

```
편의점수[편의점수['생존률'].isnull()].head()
```

	상권코드	상권코드명	일반편의점점포수	프랜차이즈점포수	총편의점점포수	개업점포수	폐업점포수	생존률
5	1000006	북촌로11길	0.0	0.0	0.0	0	0	NaN
8	1000009	삼청로5길	0.0	0.0	0.0	0	0	NaN
76	1000082	한강대로7길	0.0	0.0	0.0	0	0	NaN
78	1000084	한남대로20길	0.0	0.0	0.0	0	0	NaN
260	1000272	동소문로6길	0.0	0.0	0.0	0	0	NaN

```
편의점수[편의점수['생존률']==0]
```

	상권코드	상권코드명	일반편의점점포수	프랜차이즈점포수	총편의점점포수	개업점포수	폐업점포수	생존률
5	1000006	북촌로11길	0.0	0.0	0.0	0	0	0.0
8	1000009	삼청로5길	0.0	0.0	0.0	0	0	0.0
76	1000082	한강대로7길	0.0	0.0	0.0	0	0	0.0
78	1000084	한남대로20길	0.0	0.0	0.0	0	0	0.0



### 종속변수(생존율) 분석 / 결측치와 이상치 확인

생존률 == NaN

생존률 == 0

생존률 == 1



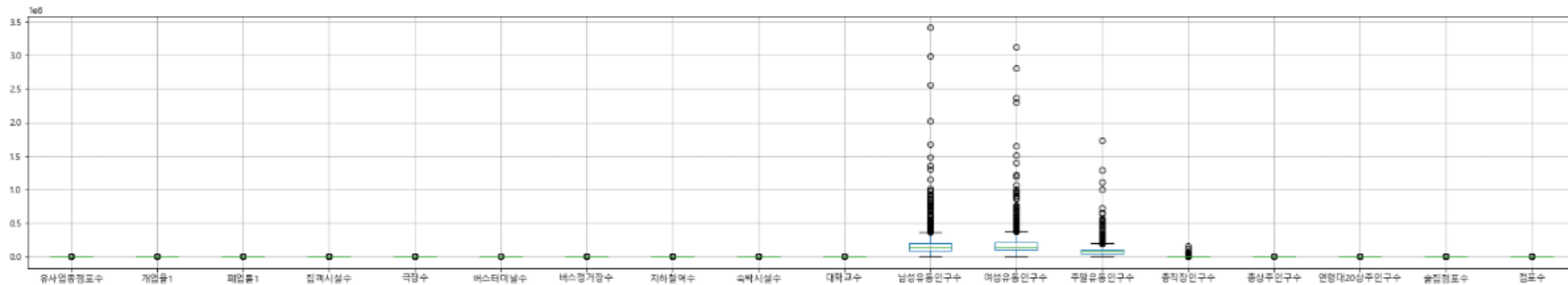
처음부터 편의점이 없는 경우

과거 편의점이 있었지만 현재 없는 경우

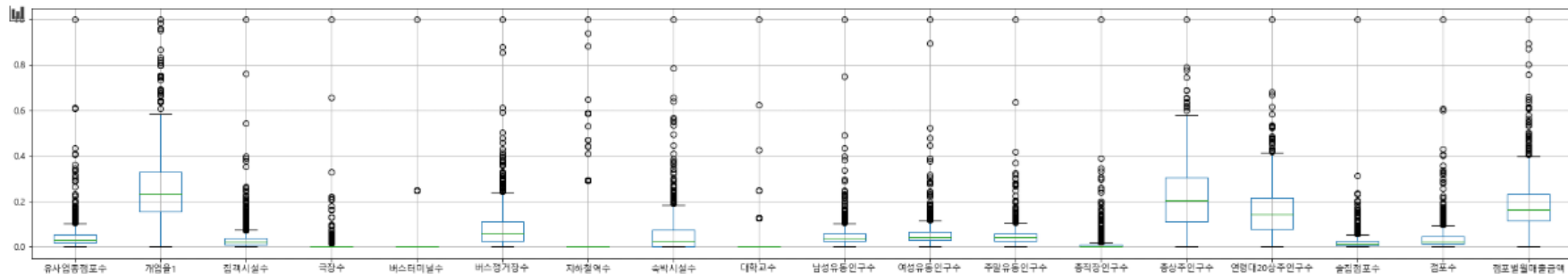
폐업점이 없는 경우



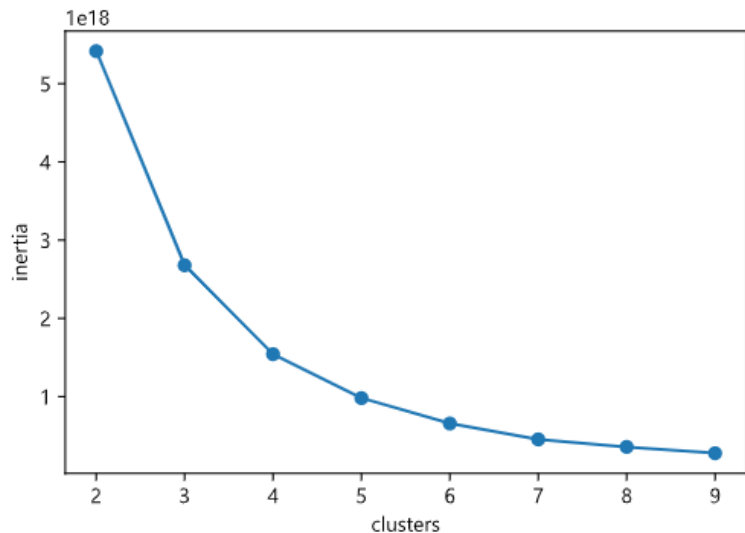
## II. 데이터 전처리



인구수가 다른 변수들에 비해 단위가 지나치게 크므로 **MinMax Scaler**을 이용하여 스케일을 조정

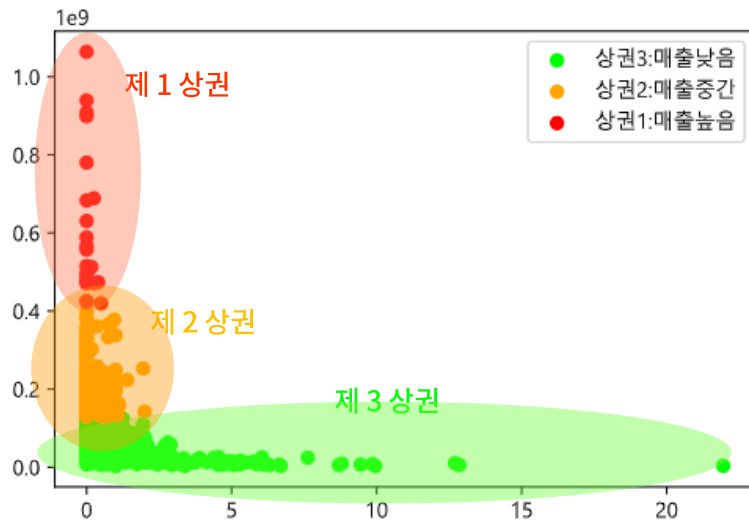


## II. 데이터 전처리



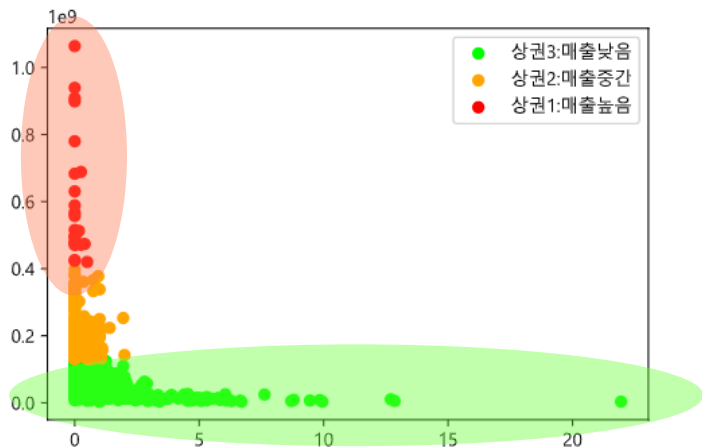
귀무가설 : 매출이 높은 상권과 낮은 상권은  
영향을 받는 독립변수가 **같다**.

대립가설 : 매출이 높은 상권과 낮은 상권은  
영향을 받는 독립변수가 **다르다**.



가설검증을 위해  
평균 매출 금액 기준  
**클러스터링**

## II. 회귀 분석



OLS Regression Results

Dep. Variable:	survive	R-squared (uncentered):	0.886
Method:	Least Squares	Adj. R-squared (uncentered):	0.762
Date:	Fri, 11 Dec 2020	F-statistic:	7.121
Time:	17:10:38	Prob (F-statistic):	0.00136
No. Observations:	23	Log-Likelihood:	-110.69
Df Residuals:	11	AIC:	245.4
Df Model:	12	BIC:	259.0
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
연령대30살주인구수	78.1476	27.540	2.837	0.016	17.529	138.766
아파트평균시가	-9.9269	41.095	-0.242	0.814	-100.377	80.524
연령대30직장인구수	95.7369	92.778	1.032	0.324	-108.466	299.940
관광서비스	-65.4033	38.562	-1.696	0.118	-150.279	19.472
은행수	-47.4740	91.172	0.521	0.613	-248.143	153.195
인구수	41.2854	29.926	1.380	0.195	-24.562	107.153
변화점수	-4.51e+13	2.34e+13	-1.928	0.080	-9.66e+13	6.4e+14
국전수	3.1957	40.325	0.079	0.938	-85.560	91.951
숙박시설수	73.2331	45.636	1.605	0.137	-27.212	173.678
의료기관	6.5822	47.743	0.138	0.893	-98.499	111.663
교육기관	7.0270	37.063	0.190	0.853	-74.548	88.602
대중교통	6.8270	39.361	0.173	0.865	-79.806	93.460
월매출평균	72.2477	37.299	1.937	0.079	-9.846	154.341

Omnibus:	0.159	Durbin-Watson:	1.689
Prob(Omnibus):	0.923	Jarque-Bera (JB):	0.052
Skew:	0.081	Prob(JB):	0.974
Kurtosis:	2.832	Cond. No.	9.87e+16

제1 상권의 경우 유의하지 않음  
(P-value > 0.05)

OLS Regression Results

Dep. Variable:	survive	R-squared (uncentered):	0.853
Method:	Least Squares	Adj. R-squared (uncentered):	0.850
Date:	Fri, 11 Dec 2020	F-statistic:	313.5
Time:	17:10:38	Prob (F-statistic):	0.00
No. Observations:	320	Log-Likelihood:	-4372.2
Df Residuals:	307	AIC:	8776
Df Model:	17	BIC:	8861
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
개업점포수	-118.0069	18.733	-6.299	0.000	-177.772	-81.242
연령대30살주인구수	85.9383	6.242	13.928	0.000	74.588	99.189
아파트평균시가	164.8740	14.574	11.313	0.000	28.273	193.475
점포수	94.4656	19.377	4.875	0.000	56.38	132.494
개업률	45.9574	14.732	3.120	0.002	17.46	74.869
연령대30직장인구수	14.1088	24.419	0.578	0.564	-33.815	62.033
관광서비스	15.4462	8.297	1.862	0.063	-9.339	31.730
은행수	3.7377	20.430	0.183	0.855	-36.29	43.833
인구수	41.0686	16.432	2.499	0.013	8.20	73.317
변화점수	-9.1527	9.217	-0.993	0.321	-27.241	8.935
국전수	-49.0065	21.191	-2.313	0.021	-90.91	-7.420
숙박시설수	33.1985	17.115	1.940	0.053	-9.391	65.788
의료기관	-7.4220	7.125	-1.042	0.298	-21.405	6.561
교육기관	50.2695	11.340	4.433	0.000	28.015	72.525
대중교통	58.3649	12.682	4.598	0.000	33.66	83.274
연령대20대주인구수	-5.9147	24.378	-0.294	0.777	-54.757	40.928
월매출평균	53.7080	3.079	17.441	0.000	46.565	59.751

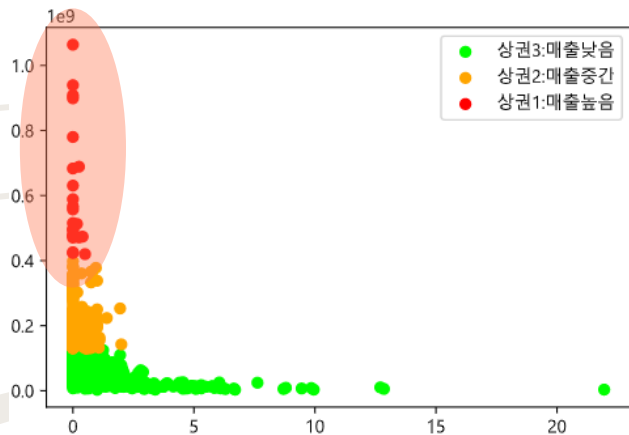
  

Omnibus:	27.888	Durbin-Watson:	1.723
Prob(Omnibus):	0.000	Jarque-Bera (JB):	54.569
Skew:	-0.169	Prob(JB):	1.41e-12
Kurtosis:	4.133	Cond. No.	18.3

제2 상권의 경우 대체적으로 유의  
(P-value < 0.05)

매출이 높은 상권은 관광지와 같은 특수상권이 다수  
독립변수에 크게 영향 받지 않을 것이라 추측

## II. 회귀 분석



제 1상권

북촌로5길  
성균관로5길  
종로58가길  
종로65길  
다산로33길

다산로36길  
동호로10길  
동호로20길  
등 23개

OLS Regression Results

Dep. Variable:	survive	R-squared (uncentered):	0.886
Model:	OLS	Adj. R-squared (uncentered):	0.762
Method:	Least Squares	F-statistic:	7.121
Date:	Fri, 11 Dec 2020	Prob (F-statistic):	0.00136
Time:	17:10:38	Log-Likelihood:	-110.69
No. Observations:	23	AIC:	245.4
Df Residuals:	11	BIC:	259.0
Df Model:	12		
Covariance Type:	nonrobust		

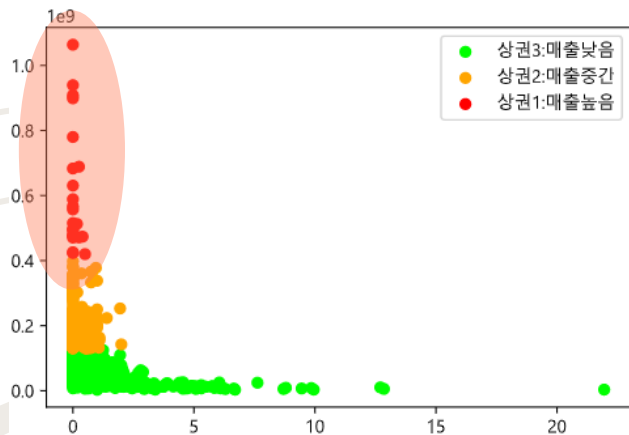
	coef	std err	t	P> t	[0.025	0.975]
연령대30상주인구수	78.1476	27.542	2.837	0.016	17.529	138.766
아파트평균시가	-9.9269	41.095	-0.242	0.814	-100.377	80.524
연령대30직장인구수	95.7369	92.778	1.032	0.324	-108.466	299.940
관공서수	-65.4033	38.562	-1.696	0.118	-150.279	19.472
은행수	-47.4740	91.172	-0.521	0.613	-148.143	153.195
약국수	41.2654	29.926	1.380	0.195	-24.582	107.153
백화점수	-4.51e-13	2.34e-13	-1.928	0.080	-1.66e-13	6.4e-14
극장수	3.1957	40.325	0.079	0.938	-85.560	91.951
숙박시설수	73.2331	45.636	1.605	0.137	-27.212	173.678
의료기관	6.5822	47.743	0.138	0.893	-98.499	111.663
교육기관	7.0270	37.063	0.190	0.853	-74.548	88.602
대중교통	6.8270	39.361	0.173	0.865	-79.806	93.460
월매출평균	72.2477	37.299	1.937	0.079	-9.846	154.341

Omnibus:	0.159	Durbin-Watson:	1.683
Prob(Omnibus):	0.923	Jarque-Bera (JB):	0.052
Skew:	0.081	Prob(JB):	0.974
Kurtosis:	2.832	Cond. No.	9.87e+16

P-value가 0.05 보다 큰 변수가 **다수**  
독립변수에 **관계없이** 매출이 높음을 확인  
실증분석을 위해 분석대상에서 **제외**

## II. 회귀 분석



제 1상권

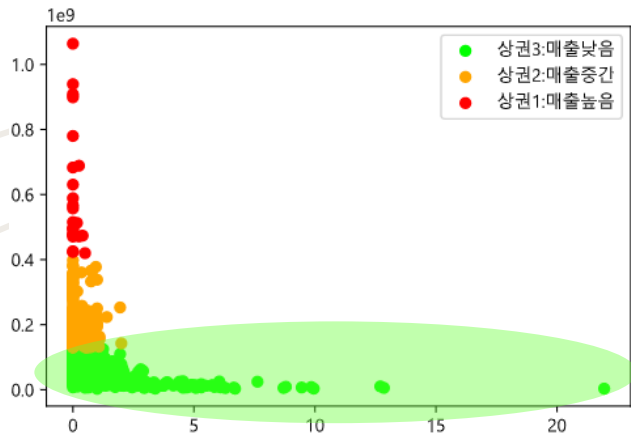
북촌로5길  
성균관로5길  
종로58가길  
종로65길  
다산로33길

다산로36길  
동호로10길  
동호로20길  
등 23개

상권코드	survive
1000022	100.000000
1000067	100.000000
1000078	100.000000
1000091	61.538462
1000264	33.333333
1000281	100.000000
1000328	100.000000
1000370	100.000000
1000440	100.000000
1000456	100.000000
1000490	100.000000
1000576	100.000000
1000592	100.000000
1000696	100.000000
1000707	45.205479
1000745	100.000000
1000756	100.000000
1000765	100.000000
1000783	60.000000
1001026	100.000000

또한 생존률의 경우 대부분의 상권이 독립변수와 상관 없이 생존율 100%

## II. 회귀 분석



제2상권

울곡로16길  
동호로11길  
퇴계로12길  
녹사평대로32길  
이태원로54길

마장로39길  
매봉길  
청계천로10나길  
등 216개

OLS Regression Results

```

=====
Dep. Variable:      survive      R-squared (uncentered):      0.830
Model:              OLS          Adj. R-squared (uncentered):    0.817
Method:              Least Squares      F-statistic:              61.19
Date:                Fri, 11 Dec 2020    Prob (F-statistic):       6.43e-68
Time:                17:10:39           Log-Likelihood:           -1054.9
No. Observations:    216              AIC:                      2162.
Df Residuals:        200              BIC:                      2216.
Covariance Type:     nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
개업점포수	-33.0155	14.180	-2.328	0.021	-60.976	-5.055
연령대30상주인구수	86.8078	12.641	6.867	0.000	61.882	111.734
아파트평균시가	116.0195	23.461	4.945	0.000	69.756	162.283
점포수	6.8404	13.122	0.521	0.603	-19.034	32.715
개업률	2.7925	26.288	0.106	0.916	-49.046	54.631
연령대30직장인구수	-13.0419	28.559	-0.457	0.648	-69.357	43.273
관공서수	22.2360	12.425	1.790	0.075	-2.264	46.736
은행수	31.3652	22.366	1.402	0.162	-12.739	75.469
약국수	35.7244	14.161	2.523	0.012	7.800	63.649
백화점수	2.1285	5.6545	0.438	0.675	-1.215	7.898
극장수	2.0278	24.499	0.083	0.934	-46.282	50.337
숙박시설수	63.4178	22.295	2.844	0.005	19.454	107.382
의료기관	-30.2993	18.948	-1.599	0.111	-67.663	7.065
교육기관	21.7413	16.687	1.303	0.194	-11.164	54.647
대중교통	20.8212	15.106	1.378	0.170	-8.966	50.608
연령대20유동인구수	64.9842	20.149	3.228	0.004	13.421	116.548
월매출평균	38.5235	9.362	4.115	0.000	20.062	56.985

```

=====
Omnibus:      1.339      Durbin-Watson:      2.041
Prob(Omnibus): 0.512      Jarque-Bera (JB):      1.438
Skew:         -0.169      Prob(JB):             0.487
Kurtosis:     2.788      Cond. No.             2.62e+18
=====

```

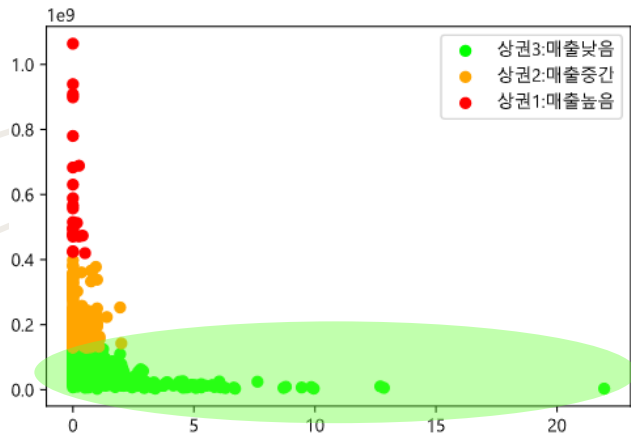
VIF Factor <= 10

	VIF Factor	features
15	3.810639	개업점포수
14	3.307748	연령대30상주인구수
13	3.104266	개업률
12	2.887485	아파트평균시가
11	2.492982	연령대20유동인구수
10	2.282770	약국수
9	2.225085	월매출평균
8	2.211211	연령대30직장인구수
7	2.208689	은행수
6	1.960490	대중교통

R2 상관계수= 83%      Durbin-Watson = 2.041

높은 설명도와 2에 가까운 D-W값이 자기 상관을 무시하고 있다는 것을 확인  
적합한 모형이라 판단

## II. 회귀 분석



제3상권

한강대로15길  
신촌로30길  
공항대로55길  
양천로24길  
디지털로37길

사당로8길  
관악로30길  
구로디지털단지  
등 938개

OLS Regression Results

Dep. Variable:	survive	R-squared (uncentered):	0.853
Model:	OLS	Adj. R-squared (uncentered):	0.850
Method:	Least Squares	F-statistic:	313.5
Date:	Fri, 11 Dec 2020	Prob (F-statistic):	0.00
Time:	17:10:39	Log-Likelihood:	-4372.2
No. Observations:	937	AIC:	8778.
Df Residuals:	920	BIC:	8861.
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
개업점포수	-118.0069	18.733	-6.299	0.000	-154.772	-81.242
연령대30상주인구수	96.9383	6.242	13.928	0.000	74.688	99.189
아파트평균시가	164.8740	14.574	11.313	0.000	136.273	193.475
점포수	94.4656	19.377	4.875	0.000	56.438	132.494
개업율	45.9574	14.732	3.120	0.002	17.046	74.869
연령대30상주인구수	14.1088	24.419	0.578	0.564	-33.815	62.033
관공서수	15.4452	8.297	1.862	0.063	-0.838	31.730
은행수	3.7377	20.430	0.183	0.855	-36.358	43.833
약국수	41.0686	16.432	2.499	0.013	8.820	73.317
병원적수	-4.1527	9.217	-0.943	0.321	-27.241	8.946
극장수	-49.0055	21.190	-2.313	0.021	-90.591	-7.420
숙박시설수	33.1985	17.115	1.940	0.053	-0.391	66.788
아동기관	-7.4220	7.125	-1.042	0.298	-21.405	6.561
교육기관	50.2695	11.340	4.433	0.000	28.015	72.525
대중교통	58.3649	12.692	4.598	0.000	33.456	83.274
월매출평균	53.7080	3.079	17.441	0.000	47.665	59.751

Omnibus:	27.888	Durbin-Watson:	1.723
Prob(Omnibus):	0.000	Jackknife (SE):	64.603
Skew:	-0.169	Prob(JB):	1.41e-12
Kurtosis:	4.133	Cond. No.	18.3

VIF Factor <= 10

	VIF Factor	features
16	8.123556	개업점포수
15	5.200558	은행수
14	3.947164	약국수
13	3.944030	연령대20유동인구수
12	3.844154	점포수
11	3.479832	숙박시설수
10	3.265832	대중교통
9	3.203327	개업율
8	3.198760	연령대30상주인구수
7	2.869450	월매출평균

R2 상관계수= 85%      Durbin-Watson = 1.723

제3 상권 분석결과 역시 적합한 모형이라 판단



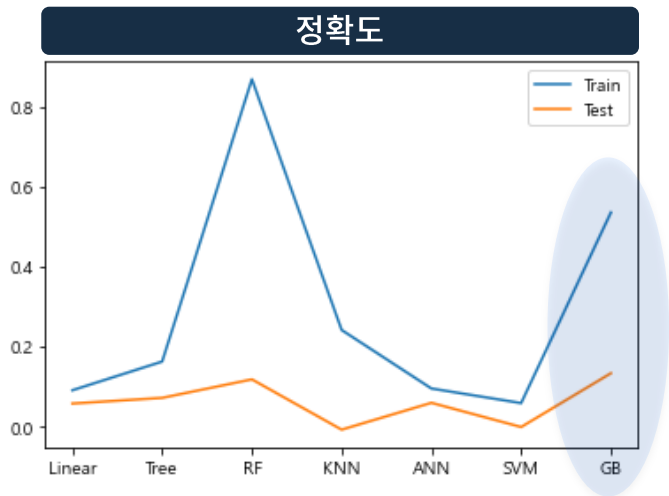
분석 결과

· 머신러닝

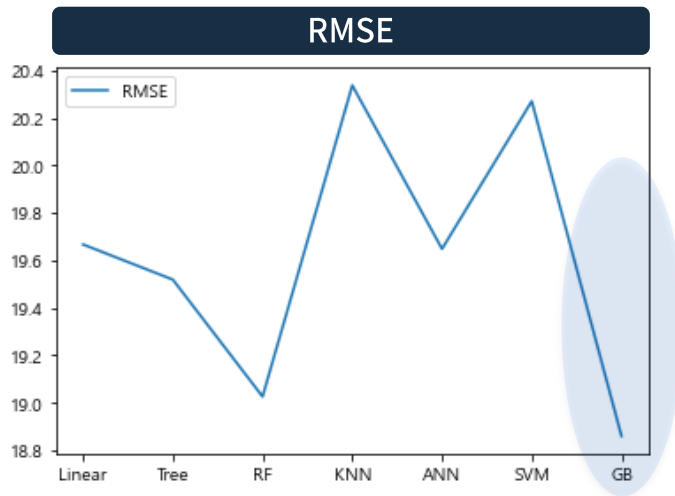


### III. 머신 러닝

- 모델 성능 비교분석: Linear Regression / Decision Tree / Random Forest / K-Means/Artificial Neural Network / SVR / Gradientboosting



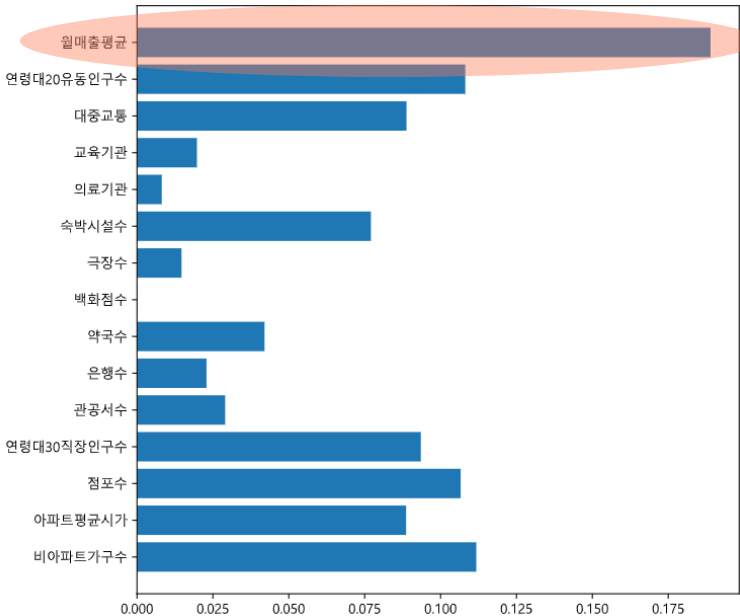
예측 정확도 면에서 개선 필요



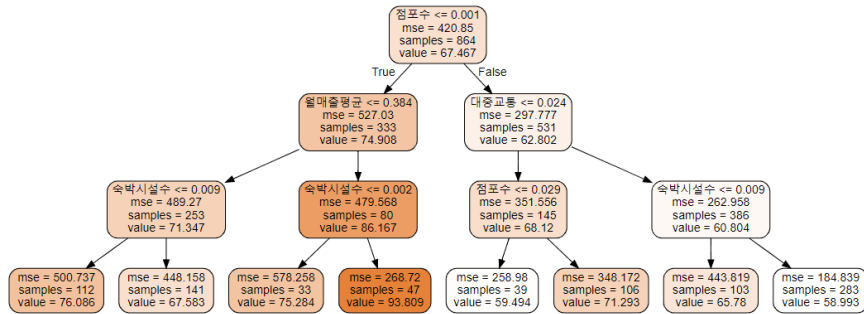
Gradientboosting 모델  
가장 낮은 오차율 18%

# III. 머신 러닝

변수 중요도



의사결정나무





## 결론

- 생존분석
- 입지추천
- 한계점 및 개선 내용

P-value < 0.05

### 양의 상관관계

비아파트 가구수 ↑

아파트평균시가 ↑

교육기관 ↑

대중교통 ↑

월매출평균 ↑

### 음의 상관관계

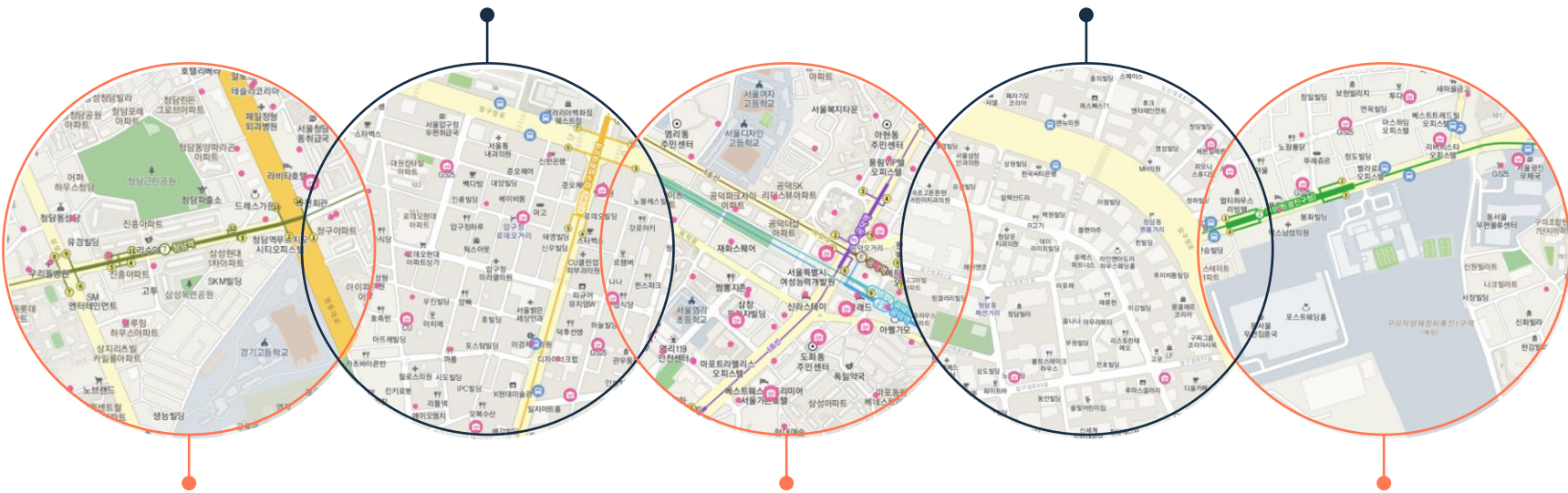
극장수 ↓

의료기관 ↓

IV. 입지추천

압구정 로데오거리

도산대로81길



청담역

공덕역

광진구 강변역

위 조건을 충족하는 최적입지 상권

## IV. 시사점 및 한계점

### 시 사 점

- ☞ 사업 추진 시 핵심 여건인 상주인구나 유동인구와 같은 인구 변수는 예상과 달리 유의하지 않음
- ☞ 서울시 매출이 높은 상위 상권은 대체로 독립변수에 영향을 받지 않고 생존율이 높음
- ☞ 편의점은 다른 업종에 비해 입점 거리 단위 기준이 작은 것을 고려해 도로명 기준으로 분석
- ☞ 편의점 개업을 목표로 하는 소상공인 대상 생존율이 높은 안정 상권 추천
- ☞ 편의점 기업이나 기존 영업중인 편의점에게는 폐업률이 높은 상권 사전 대비를 위한 정보 제공

### 한 계 점

- ☞ 도로명 단위로 존재하는 데이터로 기계학습을 위한 많은 row를 확보할 수 없음
- ☞ 또한 도로명 단위 데이터로 개별 편의점 정보를 반영하지 못함
- ☞ 편의점 규모, 횡단보도, 지출정보 등과 같은 다양하고 세부적인 환경변수를 반영하지 못함
- ☞ 코로나-19나 주식과 같은 경제적인 상황을 반영해 시기적인 요소를 분석할 것



## 참고문헌

- 한국부동산학회논문 (2014) 편의점 입지 선정시 매출에 영향을 미치는 요인분석
- 이임동 부동산학연구(2010) 편의점 매출에 영향을 미치는 입지요인에 대한 연구
- 서울시 빅데이터 공모전(2019) 카페 최적 입지 선정을 위한 조건