

카페 창업 어디서 할까?

서울시 공공데이터를 이용한 동별 카페 매장 수 분석



[계획단계] 분석 기본 가정과 계획

1. 분석대상: 걱정 카페 개수

2. 기본 가정:

- a. (면적 당) 유동인구, 여성비율, 회사 밀집 등 변수에 따라 걱정 카페 개수가 있을 것이다.
- b. 걱정 카페 개수에 비해 카페가 적은 지역일 수록 창업하기 좋은 지역일 것이다.

3. 계획:

- a. 예상 feature들과 동 별 면적을 수집, 면적 당 feature들의 값을 구한다
- b. 걱정 카페 개수를 도출하기 위한 feature들을 확정한다
- c. 확정된 feature들을 이용, 다중회귀분석을 통해 걱정 카페 개수를 도출한다

[계획단계] 예상 feature 설정[면적 당]

1. 사업체 종사자 수
2. 여성 종사자 수
3. 식당 수
4. 주점 수
5. 학원 수
6. 대학생 수
7. 게임방 수
8. 편의점 수
9. 중식당 수

[수집 정제단계] 데이터 수집

1. 생계형사업 분포현황(사업체수 기준) - data.seoul.go.kr
2. 사업체현황 종사자수(산업대분류별/동별/성별) - data.seoul.go.kr
3. 서울시 우리마을가게 상권분석서비스 - data.seoul.go.kr
4. 인구밀도(동별) - data.seoul.go.kr

[수집 정제단계] 데이터 정제 기준 설정

1. 행정동 정제: 423개 동 사용(현재 서울시 동 개수 426개)
 - a. 2018년 1월부로 동 전체 재건축에 들어간 둔촌 1동을 분석대상에서 제거
 - b. 상일1동과 상일2동이 나뉘어있다면 병합(2021년에 분할됨)
 - c. 향동이 있다면 제거(2020년에 신설)
2. 데이터 정렬
 - a. 년도 ~ 자치구 ~ 행정동 순서로 정렬(모두 오름차순)
3. 수집 데이터 기간: 2015 ~ 2019

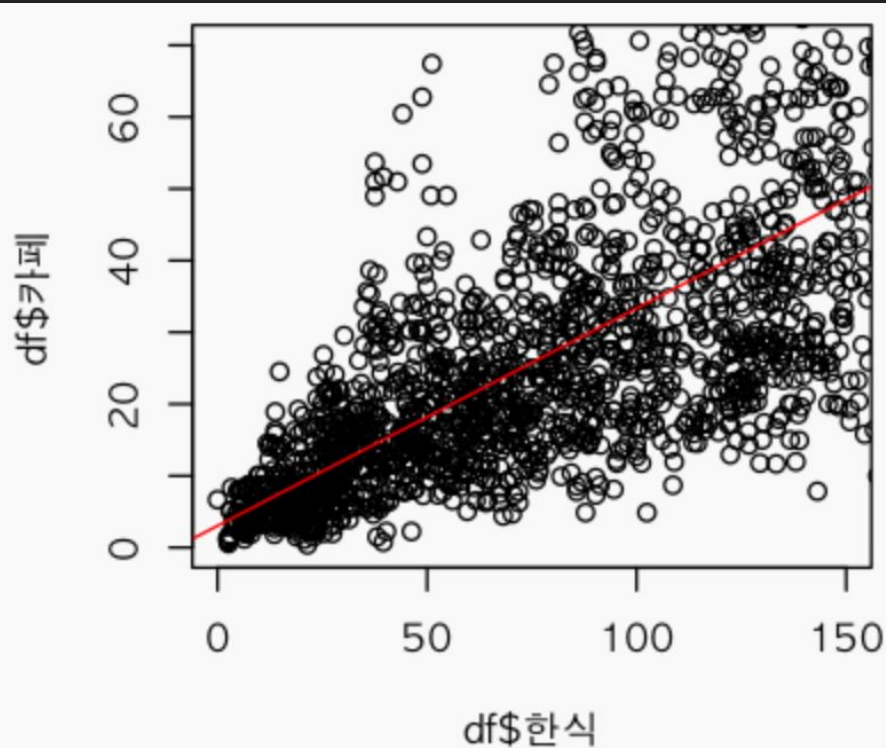
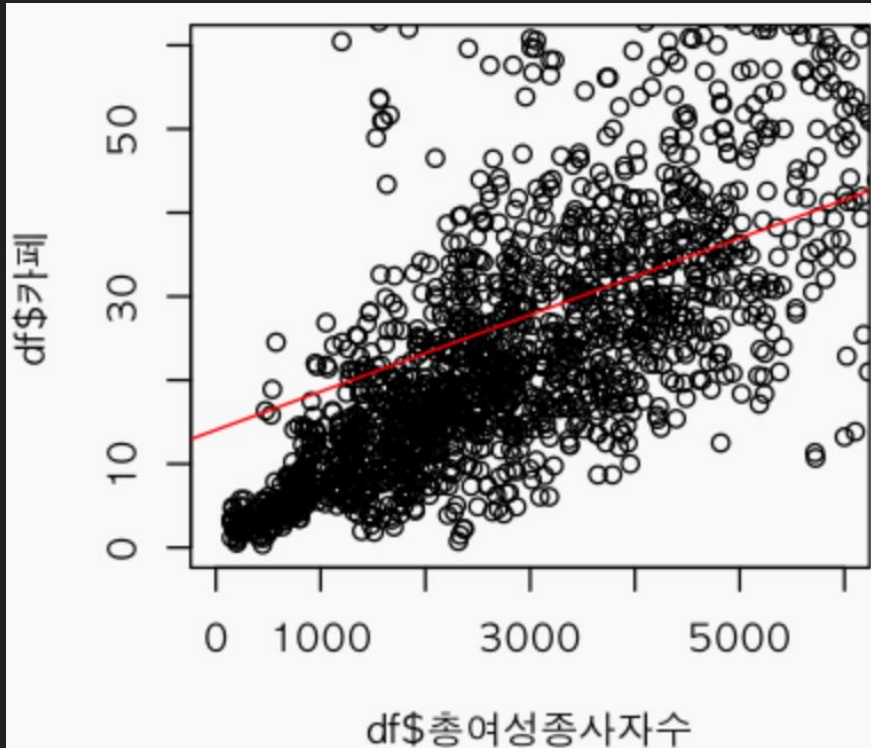
target 정제

기간	자치구	동	한식음식점업	용달 및 개별화물자동차운송업	부동산자문 및 중개업	의료점	기타주점업	두발미용업	슈퍼마켓	...	노래연습장 운영업	체인화의점	치킨전문점	자동차전문수리업	과실 및 채소매업	육류소매업	중국음식점업	당구장 운영업	제과점업
2015	합계	합계	47025	36249	21730	23683	19039	16821	10617	...	6220	7028	4471	3947	3673	3597	4002	2457	3923
2015	종로구	소계	2474	614	537	706	728	261	488	...	127	252	119	43	73	66	175	89	149
2015	종로구	사직동	334	17	56	45	91	22	34	...	12	28	12	5	5	2	22	10	20
2015	종로구	삼청동	64	8	2	59	6	1	16	...	-	9	-	-	-	-	3	1	10
2015	종로구	부암동	27	13	14	14	12	9	9	...	2	9	2	8	-	-	5	1	3
2015	종로구	평창동	60	17	31	15	27	17	13	...	2	6	3	4	2	6	6	2	7
2015	종로구	무악동	12	329	10	4	3	10	5	...	2	4	4	2	2	2	2	-	2
2015	종로구	교남동	28	13	15	8	4	5	6	...	3	6	4	1	-	-	2	1	2
2015	종로구	가회동	75	10	14	22	14	9	10	...	3	7	3	2	2	3	6	2	8

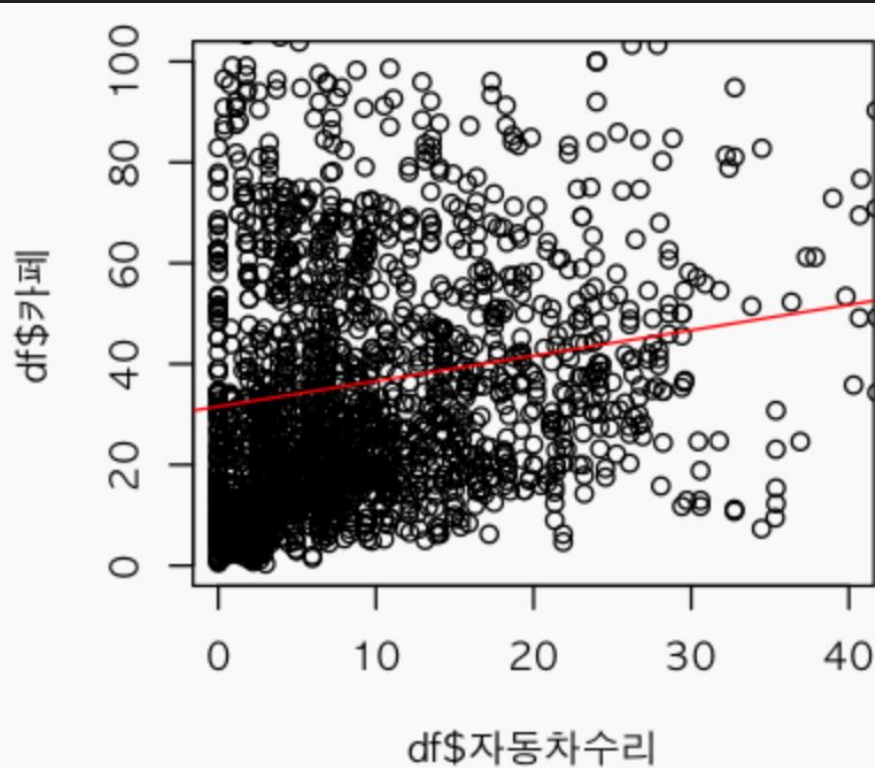
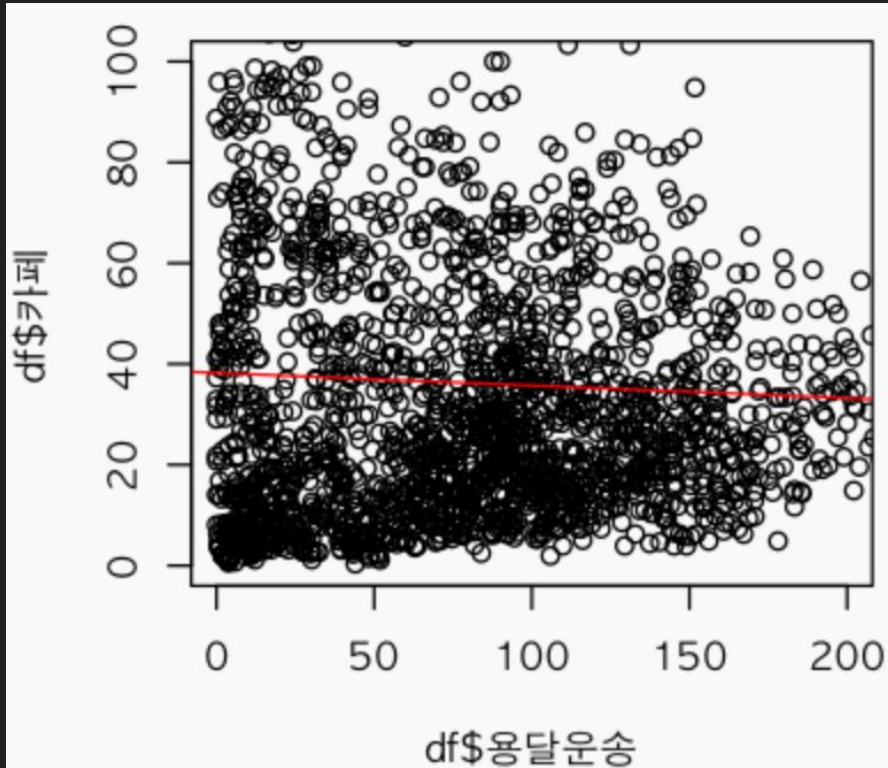


기간	자치구	동	음료점
2015	강남구	개포1동	4
2015	강남구	개포2동	14
2015	강남구	개포4동	31
2015	강남구	논현1동	103
2015	강남구	논현2동	128
2015	강남구	대치1동	25
2015	강남구	대치2동	87
2015	강남구	대치4동	102
2015	강남구	도곡1동	42
2015	강남구	도곡2동	40
2015	강남구	삼성1동	114
2015	강남구	삼성2동	90
2015	강남구	세곡동	12
2015	강남구	수서동	22
2015	강남구	신사동	140

[분석단계] plot을 그리며 사용할 feature 선정



[분석단계] plot을 그리며 사용할 feature 선정



[분석단계] 보조 지표로 correlation 확인

```
In [47]: cor(ft2015$총종사자수, ft2015$카페)
```

```
0.990314799350808
```

```
In [57]: cor(ft2016$총종사자수, ft2016$카페)
```

```
0.984163075267662
```

```
In [58]: cor(ft2017$총종사자수, ft2017$카페)
```

```
0.987516617853167
```

```
In [59]: cor(ft2018$총종사자수, ft2018$카페)
```

```
0.98896302112304
```

```
In [60]: cor(ft2019$총종사자수, ft2019$카페)
```

```
0.985995248417277
```

[분석단계] 최종 선택된 features

```
reduced <- data.frame(카페=df$카페, 한식=df$한식,  
편의점=df$편의점, 제과점=df$제과점, 부동산=df$부동산,  
전체=df$총종사자수, 전체2=df$총여성종사자수)
```

[분석단계] 분석 결과


```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.09791    0.02027  -4.830 1.92e-06 ***
한식         0.16571    0.03725   4.448 1.11e-05 ***
편의점       0.19576    0.04623   4.234 2.82e-05 ***
제과점       0.34754    0.02941  11.815 < 2e-16 ***
부동산      -0.15811    0.03116  -5.074 5.88e-07 ***
전체종사자   0.26025    0.12160   2.140  0.0329 *
여성종사자   0.17457    0.12610   1.384  0.1670

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.39 on 416 degrees of freedom
Multiple R-squared:  0.8225,    Adjusted R-squared:  0.8199
F-statistic: 321.3 on 6 and 416 DF,  p-value: < 2.2e-16
```

[분석단계] 결론

마포구 서교동

A photograph of two people riding bicycles away from the camera on a dirt path. The person on the left is wearing a dark t-shirt and light shorts, carrying a bag. The person on the right is wearing a light-colored long-sleeved shirt and patterned pants. The path is flanked by dry grass and shrubs, with a dense forest in the background. The lighting is warm and golden, suggesting late afternoon or early morning.

최선을 다한 만큼 아름답다