

[1단계] 지역선택

☒ 지번 ☐ 도로명

주소

상권분석
안하고
창업하신다고요?

초기화

분석하기

분석하기 버튼

데이터 기반 상권분석 서비스

[5조]

강은정, 정수영, 정형수, 이소현



스마트인재개발원
Smart Human Resources Development

목차

1. 문제정의

2. 데이터 선정 및 수집

3. 데이터 분석 및 시각화

3. 서비스 제안 및 결과

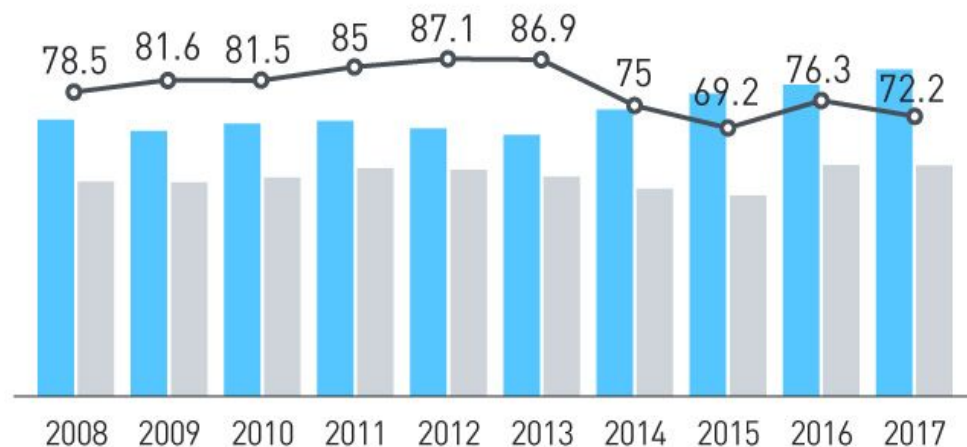
4. 팀원소개 & 질의응답

1. 문제정의 : 준비 미흡으로 인한 신규 대비 높은 폐업률

신규 대비 폐업 비율 (단위:건,%)



■ 신규 ■ 폐업
○ 신규 대비 폐업률



자료 국세청 인포그래픽 강준희

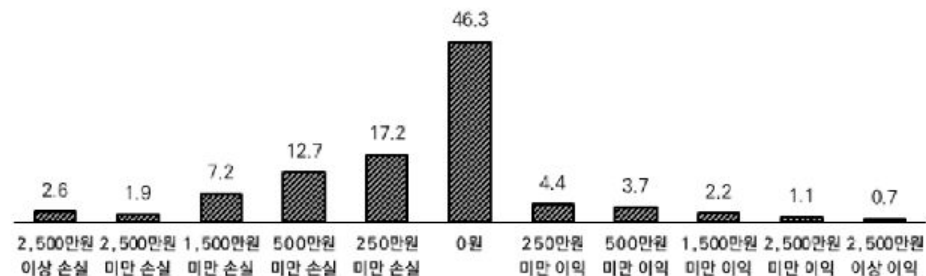
KBS

○ 현재 월평균 순이익을 살펴보면, '250만원 미만 손실'이 17.2%로 가장 높게 나타났으며, 다음으로 '500만원 미만 손실'(12.7%), '1,500만원 미만 손실'(7.2%) 등의 순으로 나타났으며, 코로나로 인한 사회적 거리두기 강화, 영업 제한 등으로 인해 순이익이 낮은 것으로 조사됨

- 종사자 수별로 살펴보면, 인원수가 많을수록 순이익 증가 비중이 높아지고 있음
- 성장단계별로 살펴보면, 성장기는 23.0%로 순이익 증가 비중이 다소 높으며, 폐업기(4.6%)로 접어들수록 순이익 증가 비중이 감소하는 것으로 조사됨

[그림 III-115] 현재 월평균 순이익(전체)

(단위 : %)



1. 문제정의

- 어디에 점포를 내지?
- 매출은 얼마나 나올까?
- 사람들이 많이 올만한 점포일까?
- 지출은 어떻게 계산하지?
- 내가 평균적인 사람들에 비해 얼마나 더 잘 준비하고 영업해야 할까?
- 개점하고 몇 년동안 살아남을 수 있지?

소상공인
프랜차이즈 담당자
대기업 PM



- 이 동이 내 업종에서 가장 높은 평균 매출을 기대할 수 있구나!
- 이 지역은 매출 대비 경쟁이 심해서 충분히 고려한 후에 개점해야겠다.
- 유동인구가 개점 업종의 매출에 이러한 영향을 주는구나!



2. 데이터 수집 : 필요한 데이터 선정



- 음식점업(카페)
- 교육업(학원)
- 소·도매업(편의점)
- 업종별 월평균 매출
- 일일 평균 유동인구
- 동종업계 경쟁 점포 수
- 업종별 매출액 및 영업이익 비율

2. 데이터 수집 : 크롤링

1. 크롤링

- 출처 : 소상공인시장진흥공단 홈페이지
- 기준 시점 : '22. 9. 28.
- 총 데이터 수 : 6,816건/약 500,000건
- 데이터 내용 : 월 평균 추정매출,

일일 평균 유동인구,

경쟁 업체 수

→ 정부기관이 제공하는 신뢰성 높은 데이터로
평균 통계치를 확인하기 위해 수집

```

31 driver.find_element_by_css_selector('#container > div.custom_zoomcontrol > div:nth-cl
32 driver.implicitly_wait(5)
33 time.sleep(1)
34 driver.find_element_by_css_selector('#container > div.custom_zoomcontrol > div:nth-cl
35 driver.implicitly_wait(5)
36 time.sleep(1)
37
38 #간단분석 클릭하기
39 driver.find_element_by_xpath('//*[@id="data_pop"]/div[2]/a').click()
40 driver.implicitly_wait(5)
41
42 #동단위로 검색
43 for k in range(len(dong_list[0])):
44     try:
45         address_search=driver.find_element_by_xpath('//*[@id="searchAddress"]')
46         driver.implicitly_wait(5)
47         time.sleep(1)
48         element = soup.select('div.cell>strong')
49         WebDriverElement = driver.find_element_by_id("address")
50         address_text = driver.get_attribute("text")
51         input_address = driver.get_attribute("value")
52         print (f'{k}번 {input_address} 을 skip합니다')
53         except:
54             pass
55
56 driver.implicitly_wait(5)
57 time.sleep(1)
58 element = soup.select('div.cell>strong')
59 WebDriverElement = driver.find_element_by_id("address")
60 address_text = driver.get_attribute("text")

```


2. 데이터 수집 : 공공데이터 및 보고서 이용



2. 중소 벤처기업부 통계청

- 출처 : 소상공인실태조사 결과 보고서
- 기준 시점 : '21.12.28.
- 데이터 내용

: 경영현황 (사업체 별 매출액, 영업이익)



$$\text{평균지출율} = (1 - \text{영업이익률})$$

$$* \text{영업이익률} = \text{연 평균 영업이익} / \text{연 평균 매출}$$

3. 데이터 분석 및 시각화

데이터 분석 전 정제

```
1 # 데이터 불러오기
2 chicken = pd.read_excel("치킨.xlsx", index_col="동")
3
4 # 불필요한 문자 삭제 및 타입 변경
5 chicken["매출"] = chicken["매출"].str.replace(",","").str.replace("만원",
6
7 # 데이터 타입 확인
8 chicken.info()
9
10 # 필요한 데이터 추출
11 chicken_re = chicken[["시장 규모","창업 후 평균 매출", "창업 후 예상 영업이익"]]
```

불필요한 문자 제거 / 데이터 타입 통일 / 필요한 데이터 추출

```
import pandas as pd
```

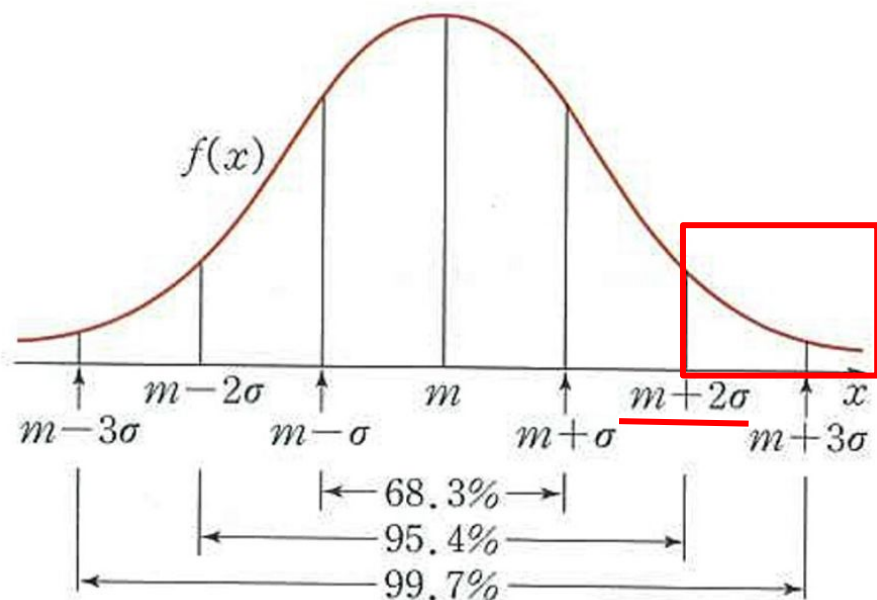
```
서울시_행정동 = pd.read_excel("서울시_행정동.xlsx")
cafe = cafe = pd.read_csv("카페.csv", index_col="동")
```

```
new = []
for name in 서울시_행정동:
    if name in cafe.index.to_list():
        pass
    else: # 누락된 행정동이 있는지 확인
        new.append(name)
```

크롤링 중 타지역 데이터가 들어왔는지 확인

3. 데이터 분석 및 시각화

```
1 # 상관관계 확인
2 academic_re.corr()
3 academic_re.describe()
4
5 # 매출 상위 2% 업체 (N+20)
6 top_2per = academic_re[academic_re["창업 후 평균 영업이익"] >= (academic_re.describe().iloc[1,2] + 2*academic_re.describe().iloc[2,2])]
7 top_2per.sort_values(by="창업 후 평균 영업이익", ascending=False)
```



생존률

영업이익

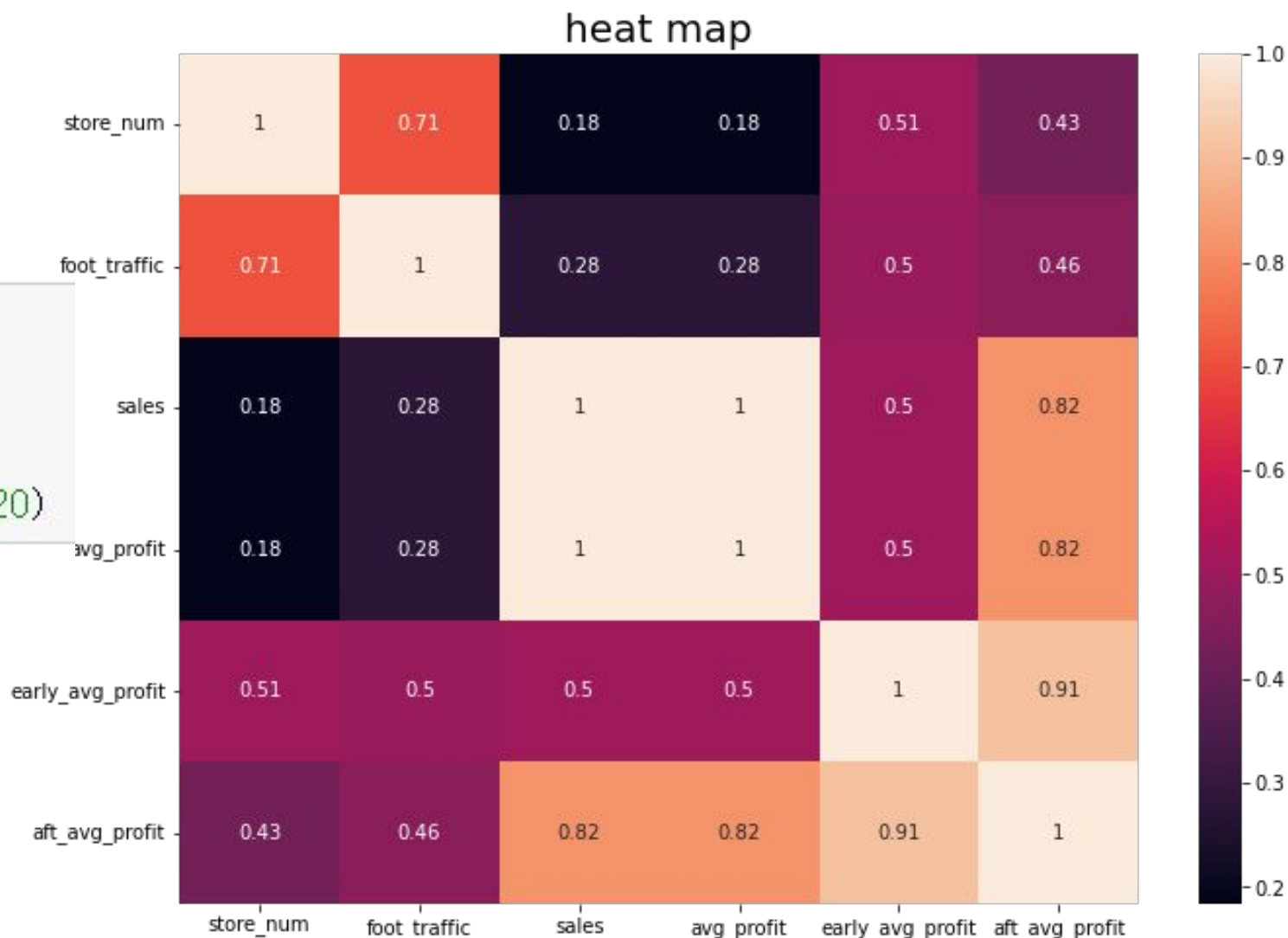
$$(100 - 95.4) / 2 = 2.2\%$$

정규분포

3. 데이터 분석 및 시각화

```
In [34]: #heatmap
fig = plt.figure()
fig.set_size_inches(11, 8)
sns.heatmap(corr2, annot=True)
plt.title('heat map', fontsize=20)
```

```
Out[34]: Text(0.5, 1.0, 'heat map')
```



4. 서비스 제안 및 보고



학원

편의점

동	창업 후 평균 영업이익	창업시 업체수	시장 규모	동	창업 후 평균 영업이익	창업시 업체수	시장 규모	창업시 매출 지표	유동인구
길음1동	1008	46	452070	고덕2동	9301	7	69282	9897	2733
잠실2동	997	26	252850	한남동	1082	15	269444	17963	7399
여의동	977	24	228689	이태원1동	886	18	264622	14701	2111
홍은2동	944	24	220800	삼각산동	857	8	113806	14226	1733
답십리2동	907	22	194523	상계8동	847	8	112441	14055	2390
중곡1동	745	7	50832
중곡1동	745	7	50832	가회동	491	7	57108	8158	1149
잠실7동	668	5	32588	월곡1동	491	12	97779	8148	2363
반포1동	667	18	117096	상암동	490	39	317224	8134	10598
거여1동	614	8	47908	중계1동	490	10	81342	8134	1594
서초1동	607	30	177596	합정동	489	19	154224	8117	4957
대치2동	594	81	469040						

5. 팀원 소개



정형수

- PM
- 웹 크롤링
- 자료수집



이소현

- PPT 제작
- 자료수집
- 웹 크롤링



강은정

- 발표
- 시각화
- 데이터 전처리



정수영

- 웹 크롤링
- 데이터 전처리

Q&A





감사합니다!