

# 9주차 캡스톤

발표 자료 PPT

<https://docs.google.com/presentation/d/1IzyqUhmePv-1S3WS4GaN17BvUFJ4XRk5wnTXvwMGpxc/edit?usp=sharing>

## < 용어 정리 >

유동인구 : 일정한 기간 동안 한 지역을 오가는 사람

상주인구 : 한 지역에 주소를 두고 거주하는 인구

생활인구 : 상주인구의 한계를 극복하기 위해 등장한 개념 - 주소만 두는 사람을 제외함

상권 : 고객들을 대상으로 점포(상품 및 서비스)의 세력이 미치는 범위.

상권 배후지 : 어떤 상업지역의 상업시설(판매시설)을 주로 이용하는 고객들이 거주하는 지역

유효수요 : 구매력을 수반하는 수요, 흔히 배후, 수요, 포텐셜 등의 용어로 표현하기도 한다.

## < 교수님 피드백 >

카페는 유동인구와 어떤 관계가 있고,

변수를 꼭 나열했을때 업종별로 그거를 확인한 패턴을 보이는데 있고 보이지 않는것도 있겠지만 데이터로 먼저 증명해보는 것이 관찰을 것!!

카페는 이러이러한 패턴을 보인다. 따라서 이런 변수가 중요하다고 데이터를 통해서 조언을 해줄수 있는 것.

업종별로 우리가 언급한 것들이 어떠한 패턴을 보이는데가 두 방향으로 진행해야 할 것.

가설이 적용이 되는가 데이터적인 수치로 보여줄수 있다면 추천도 자연스럽게 될 것.

<https://golmok.seoul.go.kr/commercialArea/commercialArea.do>

서울시 상권분석 서비스 참조

9주차때 보여줘야할 것 - 분석하려고 하는 데이터들 간의 상관관계.

가설: 유동인구가 많은 곳에 매출이 높을 것이고 좋은 상권일 것이다.

=> 데이터 분석, 시각화를 통해서 증명하기. 유동인구로 사용할 데이터들로 상권을 분석하는 것에 설득력이 생길 것

만약 유동인구가 많지만 매출이 그만큼 나오지 않는다면? 흐르는 입지일 가능성이 있다.

흐르는 입지는 버스정류장 앞 대로변처럼 유동인구는 많으나 정작 매출은 별로 일어나지 않는 곳이다. 유동인구의 절대량보다 유효 수요의 주동선에 위치한 입지가 좋은 상권에 속한다.

유효 수요의 주동선에 위치한다는 것: 회사 근처 편의점, 카페를 이용하는 것은 잦은

일이지만 횟수가 더 적더라도 특정 물건을 사기위해, 혹은 목적 구매. 많은 매출에 기여를 함

- 김중을 저, 나는 집대신 상가에 투자한다.

- <https://m.blog.naver.com/iam6819/222166479796> -> 유효수요, 주동선

- <https://m.blog.naver.com/iam6819/222167566918> -> 흐르는 입지, 법

1. 유동인구
2. 유효수요와 주동선
3. 흐르는 입지

어떤 데이터로 분석할지

< API 목록 >

<https://data.seoul.go.kr/dataList/OA-12252/S/1/datasetView.do>

- 지하철 호선별 역별 시간대별 승하차 인원 정보 (시간대별)

<https://data.seoul.go.kr/dataList/OA-12914/S/1/datasetView.do>

- 지하철 호선별 역별 승하차 인원 정보 (시간대별x, 일별O)

<https://data.seoul.go.kr/dataList/OA-12913/S/1/datasetView.do>

- 버스노선별 정류장별 시간대별 승하차 인원 정보

<https://data.seoul.go.kr/dataList/OA-15568/S/1/datasetView.do#AXexec>

- 상권-생활인구

<https://data.seoul.go.kr/dataList/OA-15584/S/1/datasetView.do>

- 상권-상주인구

<https://data.seoul.go.kr/dataList/OA-15580/S/1/datasetView.do>

- 상권-집객시설

<https://data.seoul.go.kr/dataList/OA-15560/S/1/datasetView.do>

- 상권영역

<https://data.seoul.go.kr/dataList/OA-15572/S/1/datasetView.do#AXexec>

- 상권-추정매출

아직 분석 기반이 확실히 마련되지 않아서 우선 어떤 데이터를 쓸 수 있을지 - 어떤 데이터가 상권을 분석하는 데 설득력이 있을지를 증명하는 것이 우선인듯.

그래서 우선 추천 부분은 보류하고, 분석에 초점을 맞추기

상권 단위로 변경한다면..? => 상권 영역 API

상권 단위로 분석을 하겠지

어떤 데이터로 어떤 분석?

- 1. 유동인구와 매출의 상관관계 분석

=> 유동인구의 절대량보다 유효 수요의 주동선에 위치한 입지가 좋은 상권이라는 것 증명 가능. (우리의 가설이 틀렸음이 증명됨.)

=> 추정매출&지하철/버스 승하차 인원 API,

=> 추정매출&집객시설 API

- 2. 유효 수요 파악

=> 유효 수요는 상품을 주기적이고 반복적으로 사줄 구매자를 의미함.

=> 추정매출&상주인구 API

- 3.

상권 단위로 분석을 하면

상권 단위의 데이터가 나오겠지

상권 단위의 데이터를 취합해서

결론 : 이 업종에는 이런 특성을 가진 상권이 적합하다?

유효수요(= 구매자)를 알아야 해당 상가의 매출을 예상할 수 있다. 그러니까 상권 추정 매출 API, 상주인구 API 연관지어보기..

서비스 업종 코드 명 분류!

< 나영 >

일식음식점

양식음식점

분식전문점

한식음식점

중식음식점 - 리스트 안보고 한식일식중식양식해버려서 중식 내가 가져옴 ㅎㅎ ㄹㄹ

< 수빈 >

패스트푸드점

커피-음료

치킨전문점

호프-간이주점: 2022년 총 4327 = 1096 + 1093 + 1084 + 1054

제과점

매출 비율: sales\_ratio

- 연령대 (10대~60대 이상)
- 시간대
- 요일별 (주중/주말)
- 성별

매출 금액: sales\_amt

- 연령대 (10대~60대 이상)
- 시간대
- 요일별 (주중/주말)
- 성별

매출 건수: sales\_co

- 연령대 (10대~60대 이상)
- 시간대
- 요일별 (주중/주말)
- 성별

매출 금액, 건수 비교 -> 가격대? 알 수 있음, 어느 연령대가 소비를 더 많이 하는지 매출에 더 기여를 하는지 알 수 있음

업종별 구분 코드

#전체데이터 : 928,239  
#2022년 데이터 : 136,820

```
import requests
import json
import pandas as pd
```

```
start = 1
end = 1000
```

```
df_hansic = pd.DataFrame()
```

```
for i in range(1,137):
```

```
    url =
```

```
    f"http://openapi.seoul.go.kr:8088/534e616f50717566323148426b4942/json/VwsmTrdarSelngQq/{start}/{end}/2022/"
```

```
    response = requests.get(url)
    contents = response.text
```

```
    json_api = json.loads(contents)
```

```
    json_data = json_api["VwsmTrdarSelngQq"]["row"]
```

```
    df = pd.json_normalize(json_data)
    df.head(10)
```

```
    df_hansic = pd.concat([df_hansic, df[df["SVC_INDUTY_CD_NM"]=='한식음식점']])
```

```
    start = start+1000
    end = end + 1000
```

```

# 강남역 상권 데이터프레임 생성
df_gangnam = pd.DataFrame()

# 호프-간이주점 데이터프레임에서 강남역상권 추출
df_bar_gangnam = pd.DataFrame()
df_bar_gangnam = pd.concat([df_bar_gangnam,
df_bar[df_bar["TRDAR_CD_NM"]=="강남역"]])

# 강남역 상권에 넣기
df_gangnam = pd.concat([df_gangnam, df_bar_gangnam])

# 패스트푸드점 데이터프레임에서 강남역상권 추출
df_fastfood_gangnam = pd.DataFrame()
df_fastfood_gangnam = pd.concat([df_fastfood_gangnam,
df_fastfood[df_fastfood["TRDAR_CD_NM"]=="강남역"]])

# 강남역 상권에 넣기
df_gangnam = pd.concat([df_gangnam, df_fastfood_gangnam])

# 커피-음료 데이터프레임에서 강남역상권 추출
df_coffee_gangnam = pd.DataFrame()
df_coffee_gangnam = pd.concat([df_coffee_gangnam,
df_coffee[df_coffee["TRDAR_CD_NM"]=="강남역"]])

# 강남역 상권에 넣기
df_gangnam = pd.concat([df_gangnam, df_coffee_gangnam])

# 새 칼럼 생성 - 데이터타입 변환
df_gangnam["분기당매출금액"] = pd.to_numeric(df_gangnam["THSMON_SELNG_AMT"])
df_gangnam["점포수"] = pd.to_numeric(df_gangnam["STOR_CO"])

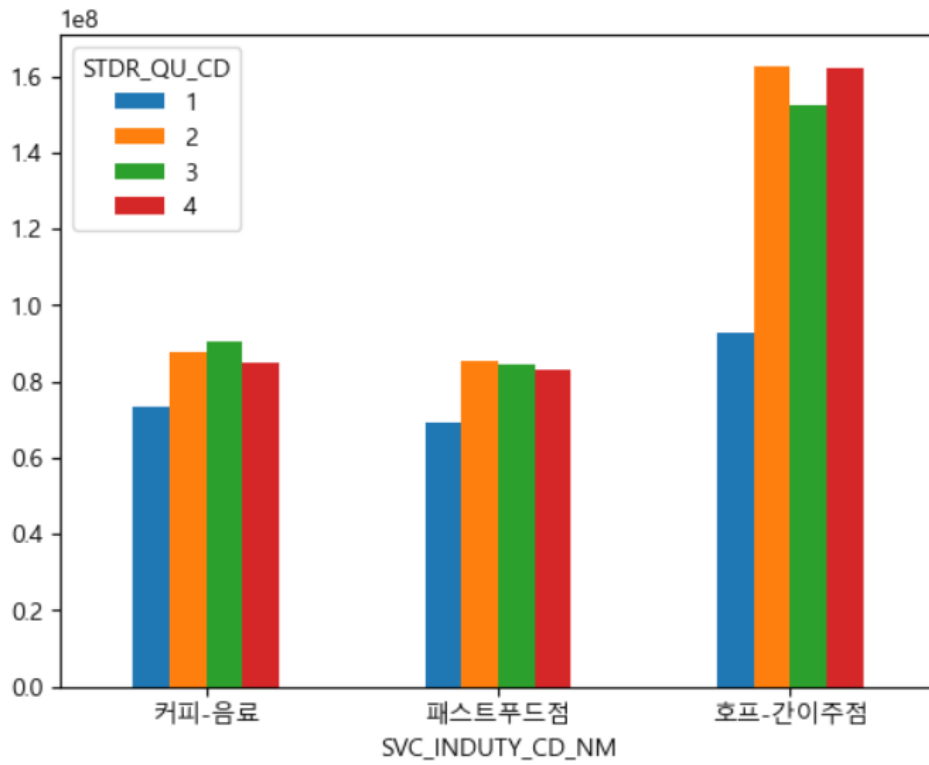
# 점포당 평균 매출금액 계산해서 새 칼럼 생성
df_gangnam["분기당매출금액/점포수"] =
df_gangnam["분기당매출금액"]/df_gangnam["점포수"]

# 업종별 1~4분기 점포 평균 매출 그래프
import matplotlib.pyplot as plt
plt.rc("font", family="Malgun Gothic")

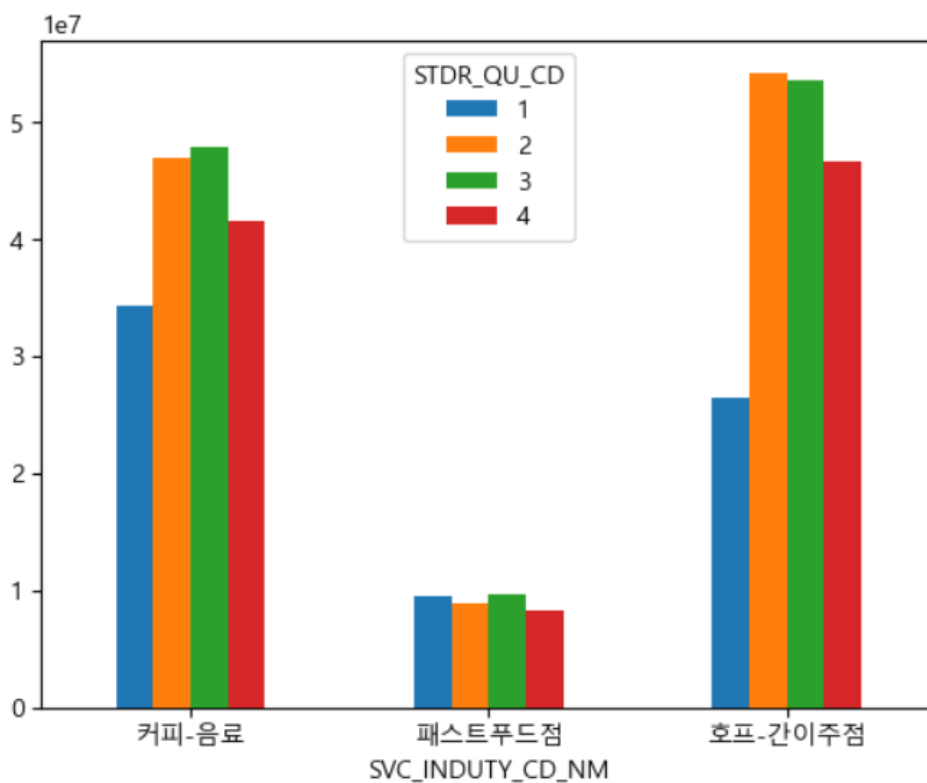
df_gangnam.pivot_table(index="SVC_INDUTY_CD_NM", columns="STDR_QU_CD",
values="분기당매출금액/점포수").plot.bar(rot=0)

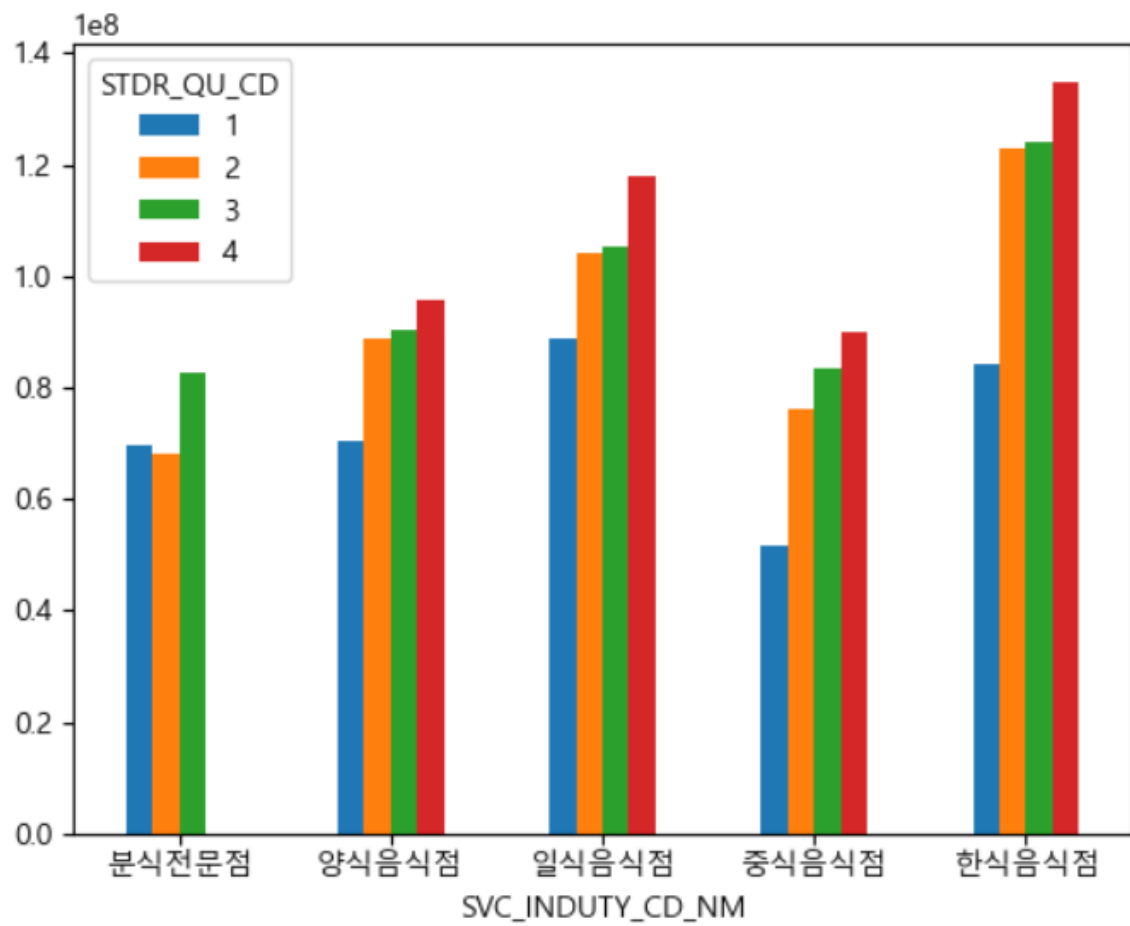
```

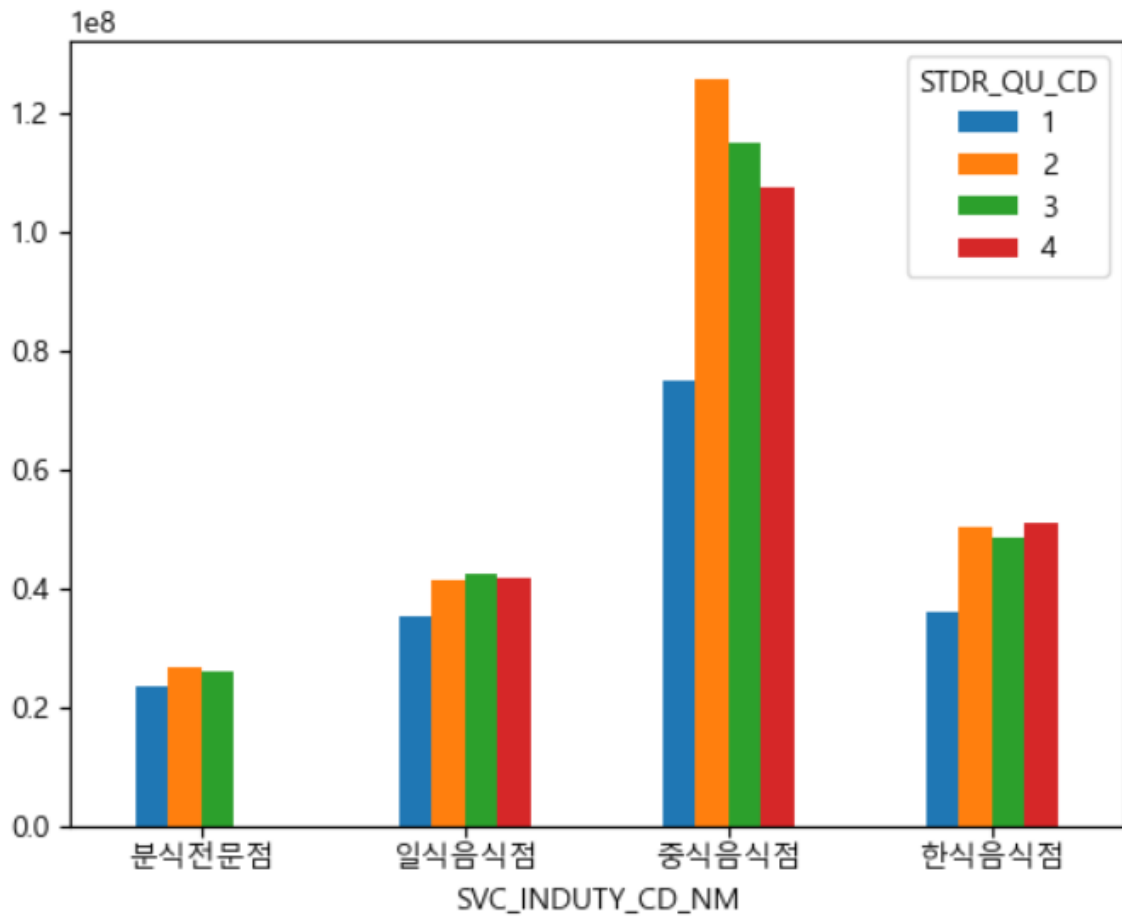
< 강남역 점포 평균 매출 >



< 창동역 점포 평균 매출 >







창동역 상권에는 양식음식점이 없음.

ㅠ=

발표할때 코드가 필요할까 싶긴 하지만,,, 일단 올려둠

지하철 승하차 유동인구



```

1 import requests
import json
import pandas as pd

url = 'http://openapi.seoul.go.kr:8088/4a7546465171'

response = requests.get(url)
contents = response.text

json_api = json.loads(contents)

json_data = json_api['CardSubwayStatsNew']['row']

df = pd.json_normalize(json_data)
df.describe()

```

	RIDE_PASGR_NUM	ALIGHT_PASGR_NUM
count	548.000000	548.000000
mean	13377.191606	13355.094891
std	11811.042254	11968.257682
min	1.000000	0.000000
25%	5093.250000	4938.250000
50%	10134.500000	9998.500000
75%	17432.500000	18031.250000
max	85882.000000	82556.000000

```

df[df['RIDE_PASGR_NUM']==85882.000000]
#최다 탑승인원

```

	USE_DT	LINE_NUM	SUB_STA_NM	RIDE_PASGR_NUM	ALIGHT_PASGR_NUM	WORK_DT
31	20230327	2호선	강남	85882.0	82556.0	20230330

```

df[df['RIDE_PASGR_NUM']==1.000000]
#최소 탑승인원

```

	USE_DT	LINE_NUM	SUB_STA_NM	RIDE_PASGR_NUM	ALIGHT_PASGR_NUM	WORK_DT
374	20230327	경원선	창동	1.0	0.0	20230330

```

df[df['ALIGHT_PASGR_NUM']==82556.000000]
#최다 하차인원

```

	USE_DT	LINE_NUM	SUB_STA_NM	RIDE_PASGR_NUM	ALIGHT_PASGR_NUM	WORK_DT
31	20230327	2호선	강남	85882.0	82556.0	20230330

```

df[df['ALIGHT_PASGR_NUM']==0.000000]
#최소 하차인원

```

	USE_DT	LINE_NUM	SUB_STA_NM	RIDE_PASGR_NUM	ALIGHT_PASGR_NUM	WORK_DT
72	20230327	3호선	충무로	2.0	0.0	20230330
374	20230327	경원선	창동	1.0	0.0	20230330
418	20230327	경의선	계양	2.0	0.0	20230330