

「Korea Polytechnic University」

**파이썬 프로그래밍 프로젝트
기말보고서**

2020-02학기

담당교수	컴퓨터공학부 서대영
학과	컴퓨터공학과
학번	2019150017
이름	박민수

파이썬 프로그래밍 프로젝트 기말보고서 목차

1 선정한 데이터와 선정 이유

2 데이터를 통해서 알고 싶은 내용들의 쿼리와 추출하는 이유

3 쿼리의 값으로 받은 결과물 해석방법

4 쿼리의 값으로 받은 결과물 해석하기

5 프로젝트를 하면서 소감

1

선정한 데이터와 선정 이유

선정한 데이터 :

소상공인시장진흥공단_상가(상권)정보_부산_20200630

(<https://www.data.go.kr/data/15012005/fileData.do>)

주민등록 인구 및 세대현황(수영구, 남구, 중구, 영도구, 동래구)

(<https://jumin.mois.go.kr/index.jsp>)

선정 이유 :

흔히들 사람들은 대학 나와서 이과 특히 공대는 회사에 취직을 했다가 치킨집을 차린다고 우스갯소리로 얘기를 한다.

그래서 만약 치킨집을 차리게 되더라도 어떤 프랜차이즈를 고르고 또 어디에 개업을 할지 고민을 많이 할 것이다. 그러한 고민들을 해결하기 위해 이 데이터를 선택하였다.

또한 프랜차이즈와 그 프랜차이즈가 있는 위치를 비교하여 고른 위치가 과연 장사가 잘 될 것인지 최근 5년간의 주거 인구를 조사하고 그 위치의 인구가 어떻게 변화하는지 조사를 하고 예측을 하기 위해서 선택하였다.

2

데이터를 통해서 알고 싶은 내용들의 쿼리와 추출하는 이유

1. 부산에 치킨 프랜차이즈 별로 몇 개의 가게가 있는지 구하라

추출 이유 : 먼저 전체적으로 어떤 프랜차이즈를 있는지 한눈에 볼 수 있고 2번의 쿼리를 수행하기 위해서 필요하다.

2. 1번에서 구한 치킨별 프랜차이즈의 개수에서 적당한 개수인 프랜차이즈를 선별하라

추출 이유 : 너무 많다면 경쟁이 심할 것이고 너무 적으면 인기가 많이 없을 것 같기 때문에 적당한 개수의 프랜차이즈를 선택하기 위해서이다.

선별 방법 : 치킨 프랜차이즈 점유율 상위 5개를 선택한다.

3. 2번에서 선정한 프랜차이즈들의 가게들을 부산광역시 지도에 표시하라

추출 이유 : 내가 선택한 프랜차이즈와 같은 가게들을 지도에 표시해서 어디에 위치해 있는지 한눈에 알고 주변에 개업을 하지 않기 위해 필요하다. 그리고 대략 어디에 개업할지 선택을 하기 위해서이다.

4. 3번에서 고른 위치의 주거인구의 변화를 조사하고 내년에는 어떻게 바뀔지 예측을 하라

추출 이유 : 3번에서 선택한 위치의 주거인구의 변화를 조사하고 내년을 예측하여 고른 위치가 적합한지 판단하기 위해서이다.

3 쿼리의 값으로 받은 결과를 해석방법

1. 부산에 치킨 프랜차이즈 별로 몇 개의 가게가 있는지 구하라

=> 각 프랜차이즈의 개수로 어느 프랜차이즈가 더 인기가 많은지 알 수 있다.

2. 1번에서 구한 치킨별 프랜차이즈의 개수에서 적당한 개수인 프랜차이즈를 선별하라

=> 부산 전체의 치킨 개수를 구하고 각 업종별로 점유율을 구하고 점유율이 상위 5개인 프랜차이즈를 선택하여 소비자에게 인기있는 프랜차이즈를 구한다.

3. 2번에서 선정한 프랜차이즈들의 가게들을 부산광역시 지도에 표시하라

=> 같은 프랜차이즈가 대략적으로 어디에 위치한지를 지도에 표시하고 상대적으로 근처에 같은 프랜차이즈가 없는 위치를 정한다.

4. 3번에서 고른 위치의 주거인구의 변화를 조사하고 내년에는 어떻게 바뀔지 예측을 하라

=> 3번에서 선택한 위치의 10년간 주거인구의 연도별 주거 인구를 구한다. 구한 값들을 최소제곱법을 이용하여 2020년의 인구수를 구하고 10년간의 평균 주거 인구보다 많으면 신장개업이 가능하다고 판단을 한다.

4 쿼리의 값으로 받은 결과를 해석하기

1. 부산에 치킨 프랜차이즈 별로 몇 개의 가게가 있는지 구하라

```
import csv
f = open('소상공인시장진흥공단_상가(상권)정보_부산_20200630.csv')
data = csv.reader(f)
header = next(data) # 열이 39개 이므로 인덱스는 0~38
chickens_count = 0
chickens_brand = {}
for row in data :
    if row[10] == "치킨 전문점": # 10열의 애트리뷰트명은 표준산업분류명
        if row[1] in chickens_brand:
            chickens_brand[row[1]] = chickens_brand[row[1]] + 1
        else:
            chickens_brand[row[1]] = 1
    chickens_count = chickens_count + 1

sorted_chickens_brand = dict(sorted(chickens_brand.items(), key=lambda x: x[1], reverse=True))
print("부산의 치킨집 총 수",chickens_count)
f.close()
```

부산의 치킨집 총 수 2322

```
import pandas as pd
chickens_brand_ranking = pd.DataFrame(list(sorted_chickens_brand.items()), columns = ["프랜차이즈", "가맹점 개수"])
chickens_brand_ranking["점유율(%)"] = [i/chickens_count * 100 for i in sorted_chickens_brand.values()]
chickens_brand_ranking
```

	프랜차이즈	가맹점 개수	점유율(%)
0	처갓집양념치킨	68	2.928510
1	교촌치킨	66	2.842377
2	BBQ치킨	50	2.153316
3	멕시칸치킨	49	2.110250
4	네네치킨	42	1.808786
...
1153	영통닭	1	0.043066
1154	닭강정반하다	1	0.043066
1155	왕모스치킨	1	0.043066
1156	오지계옛날통닭	1	0.043066
1157	신동키닭	1	0.043066

1158 rows × 3 columns

2. 1번에서 구한 치킨별 프랜차이즈의 개수에서 적당한 개수인 프랜차이즈를 선별하라

```
chickens_brand_ranking[:5]
```

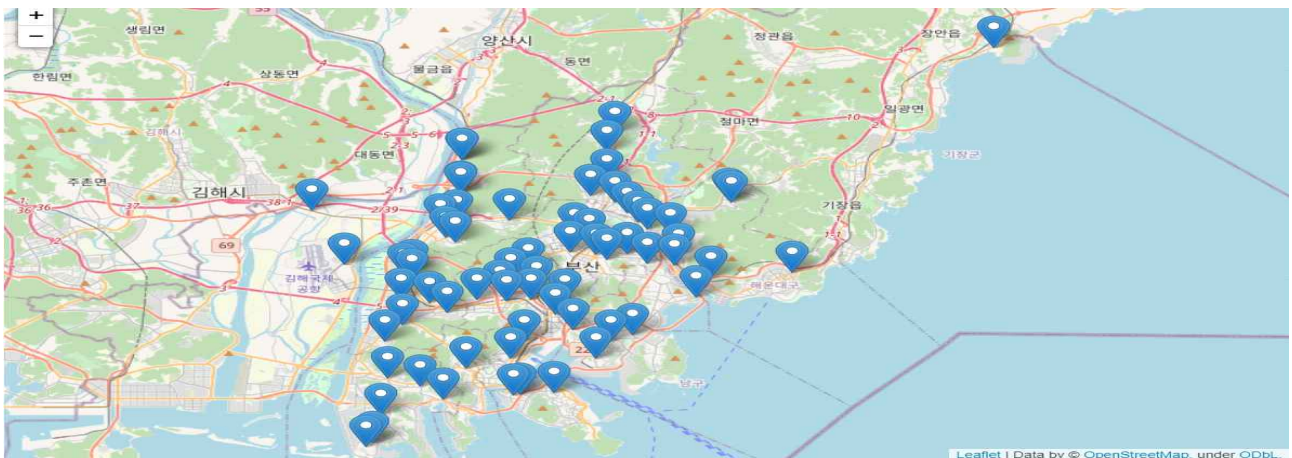
	프랜차이즈	가맹점 개수	점유율(%)
0	처갓집양념치킨	68	2.928510
1	교촌치킨	66	2.842377
2	BBQ치킨	50	2.153316
3	멕시칸치킨	49	2.110250
4	네네치킨	42	1.808786

점유율 상위 5개의 프랜차이즈를 선택한다.

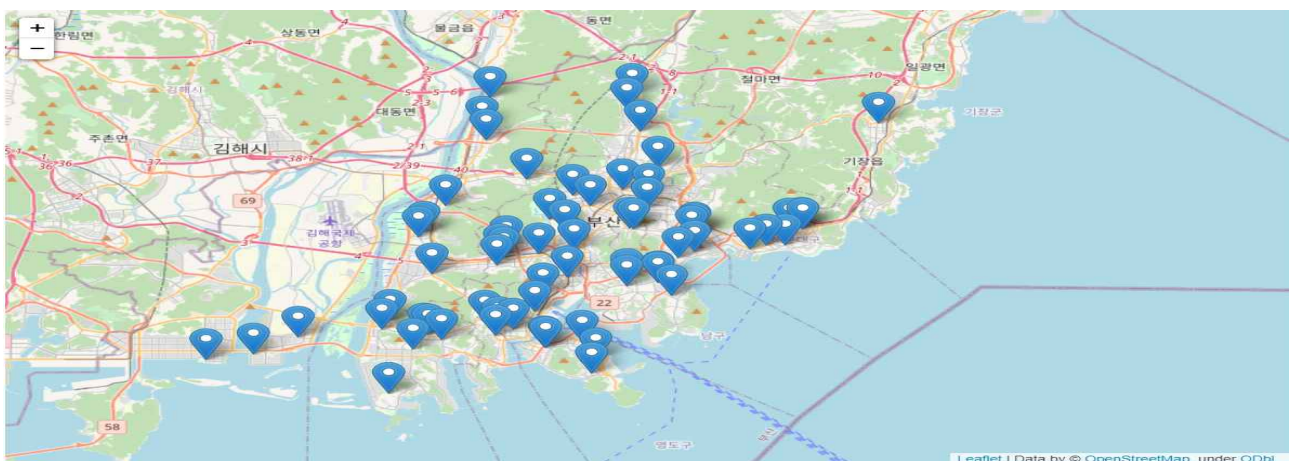
(점유율이 많다고 소비자가 많은 것은 아니지만 점유율이 많으면 소비자에게 더 인기 있다고 가정한다.)

3. 2번에서 선정한 프랜차이즈들의 가게들을 부산광역시 지도에 표시하라

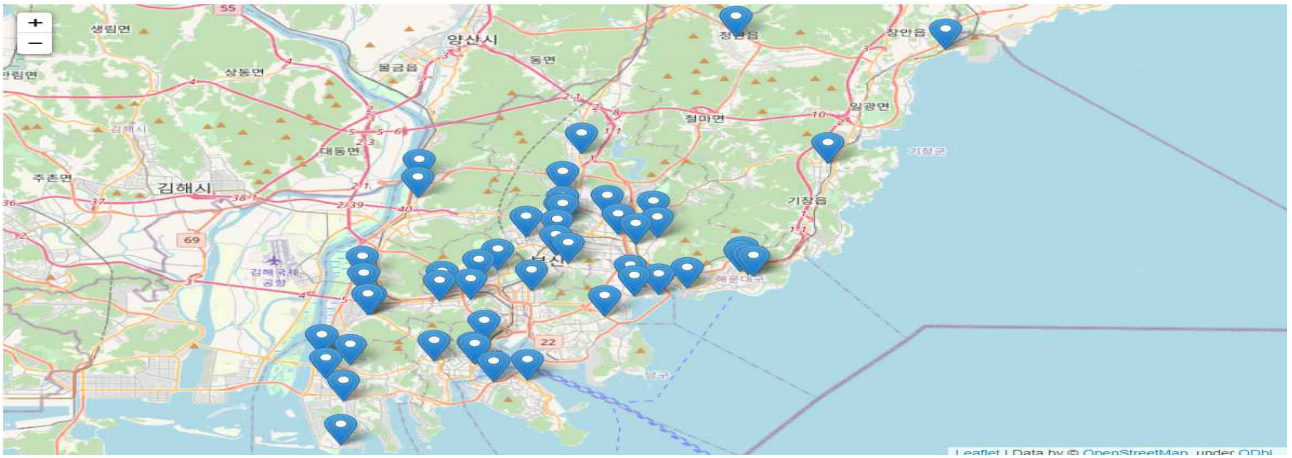
<처갓집양념치킨>



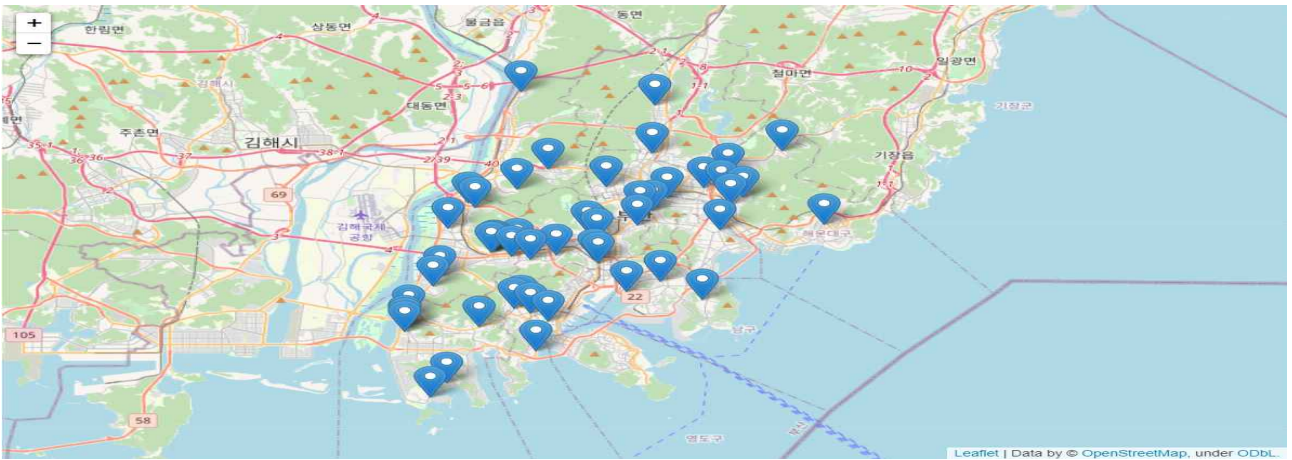
<교촌치킨>



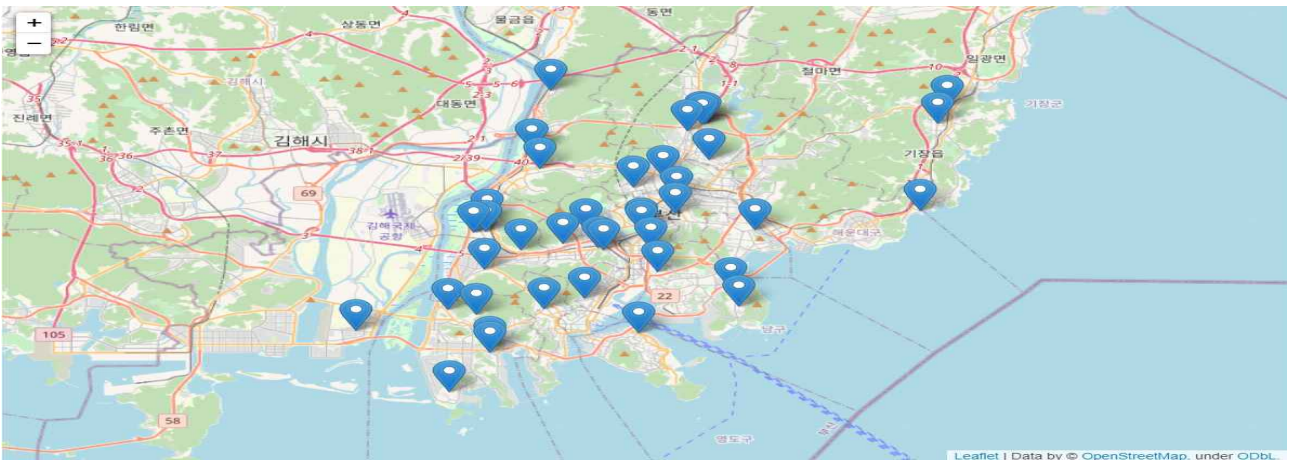
<BBQ치킨>



<멕시칸치킨>



<네네치킨>



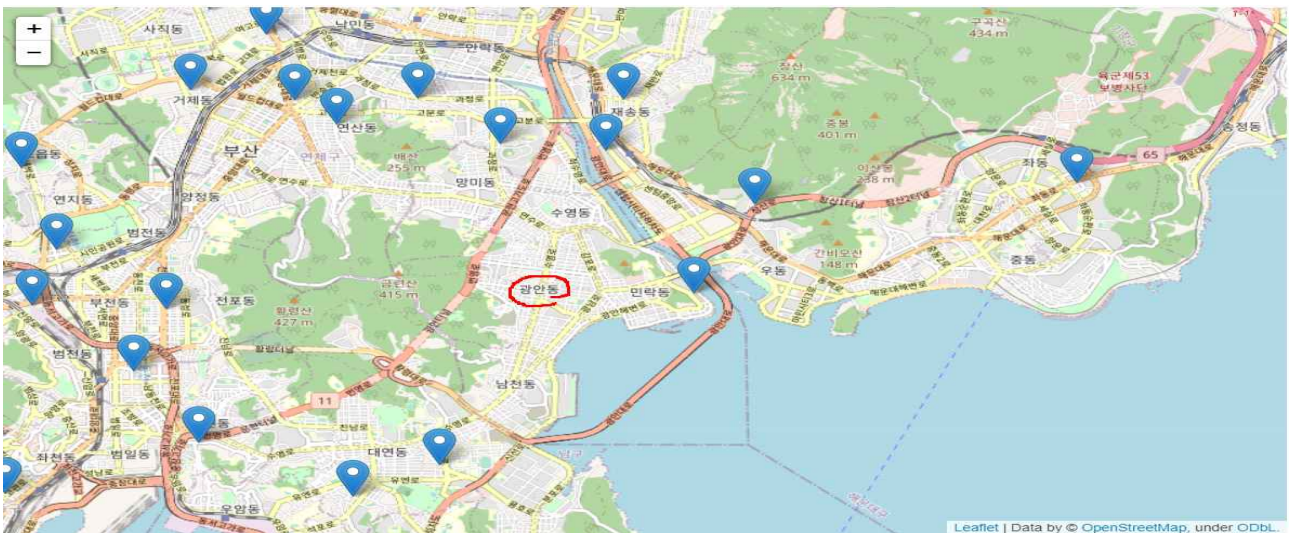
4. 3번에서 신장개업할 위치를 선정하고 주거인구의 변화를 조사하고 내년의 인구를 예측하고 신장개업 가능여부를 판단하라.

최소제곱법 : 어떤 계의 해방정식을 근사적으로 구하는 방법. $y = ax + b$

$$a = \frac{\sum_{i=0}^n (x_i - x_m)(y_i - y_m)}{\sum_{i=0}^n (x_i - x_m)^2}, b = y_m - ax_m$$

개업 위치를 선정하는 기준 : 주변에 같은 프랜차이즈가 없는 곳

< 처갓집양념치킨 - 광안동(수영구) >



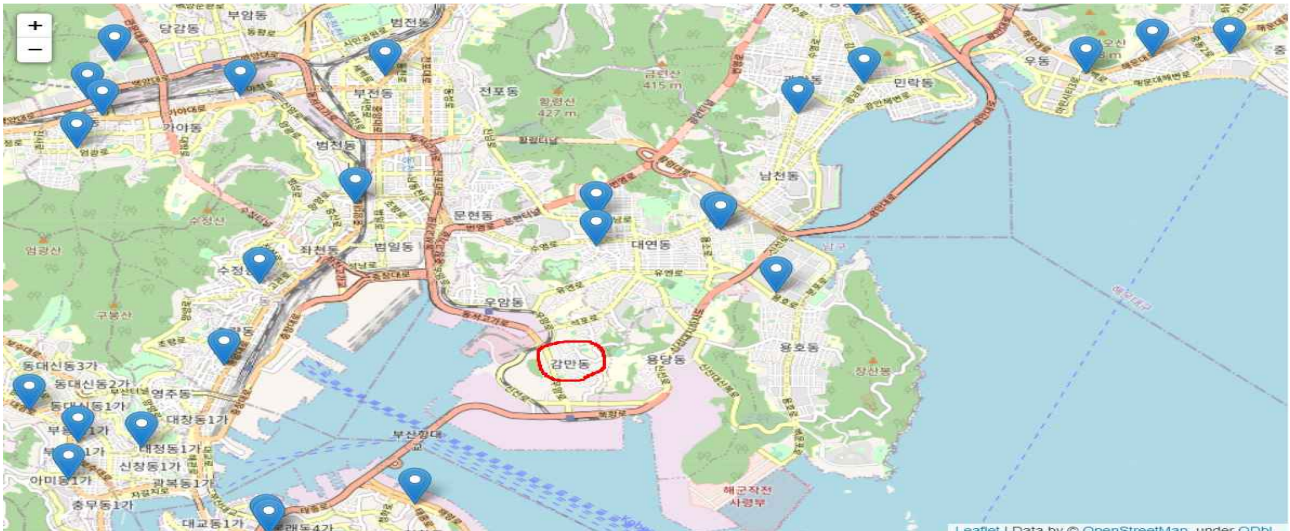
```
import csv
f = open('201012_201912_주민등록인구및세대현황_연간(수영구).csv')
data = csv.reader(f)
header = next(data)
years = [2010,2011,2012,2013,2014,2015,2016,2017,2018,2019]
suyoung_pop = [0,0,0,0,0,0,0,0,0,0]
for row in data :
    if "광안" in row[0] :
        for i in range(10):
            suyoung_pop[i] = suyoung_pop[i] + int(row[3*i+1].replace(',',''))
```

```
def top(X, mx, Y, my):
    d = 0
    for i in range(len(X)):
        d += (X[i] - mx)*(Y[i]-my)
    return d
```

```
import numpy as np
avg_year = sum(years)/10 # x의 평균
avg_suyoung_pop = sum(suyoung_pop)/10 # y의 평균
divisor = sum([(1-avg_year)**2 for i in years])
dividend = top(years, avg_year, suyoung_pop, avg_suyoung_pop)
a = dividend / divisor # 최소 제곱해의 기울기
b = avg_suyoung_pop - (avg_year * a) # 최소 제곱해의 y절편
predict_2020 = a * 2020 + b
if avg_suyoung_pop < predict_2020:
    print("광안동에서 처갓집 프랜차이즈 신장개업이 가능합니다.")
else:
    print("광안동에서 처갓집 프랜차이즈 신장개업이 불가능합니다.")
```

광안동에서 처갓집 프랜차이즈 신장개업이 가능합니다.

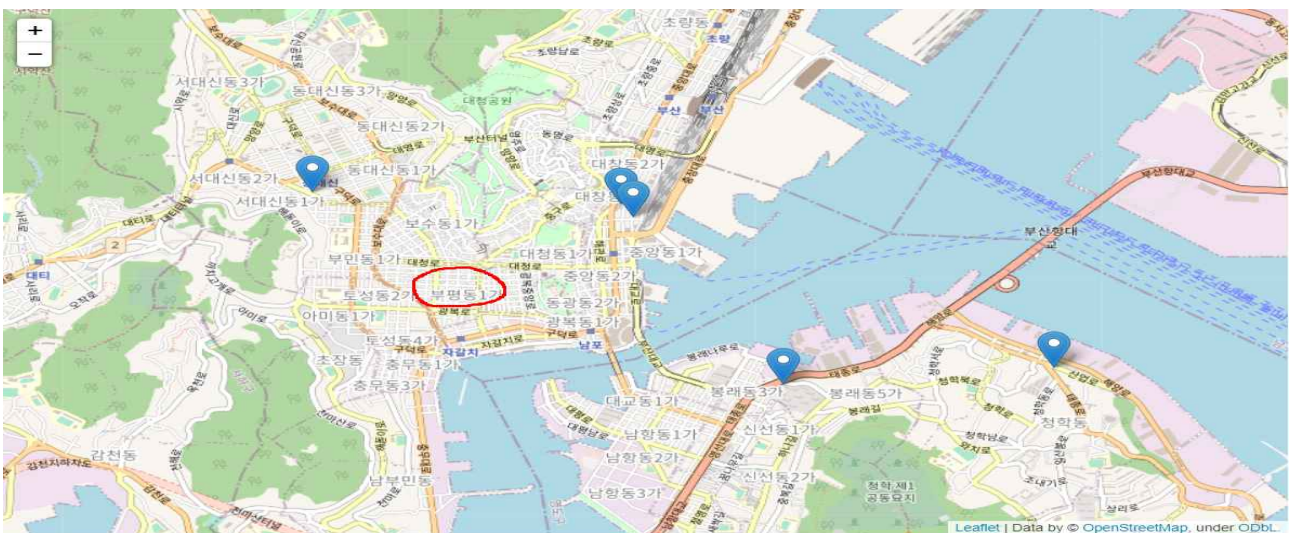
< 교촌치킨 - 감만동(남구) >



```
avg_year = sum(years)/10 # x의 평균
avg_nam_pop = sum(nam_pop)/10 # y의 평균
divisor = sum([(1-avg_year)**2 for i in years])
dividend = top(years, avg_year, nam_pop, avg_nam_pop)
a = dividend / divisor # 최소 제곱해의 기울기
b = avg_nam_pop - (avg_year * a) # 최소 제곱해의 y절편
predict_2020 = a * 2020 + b
if avg_nam_pop < predict_2020:
    print("감만동에서 교촌집 프랜차이즈 신장개업이 가능합니다.")
else:
    print("감만동에서 교촌집 프랜차이즈 신장개업이 불가능합니다.")
```

감만동에서 교촌집 프랜차이즈 신장개업이 불가능합니다.

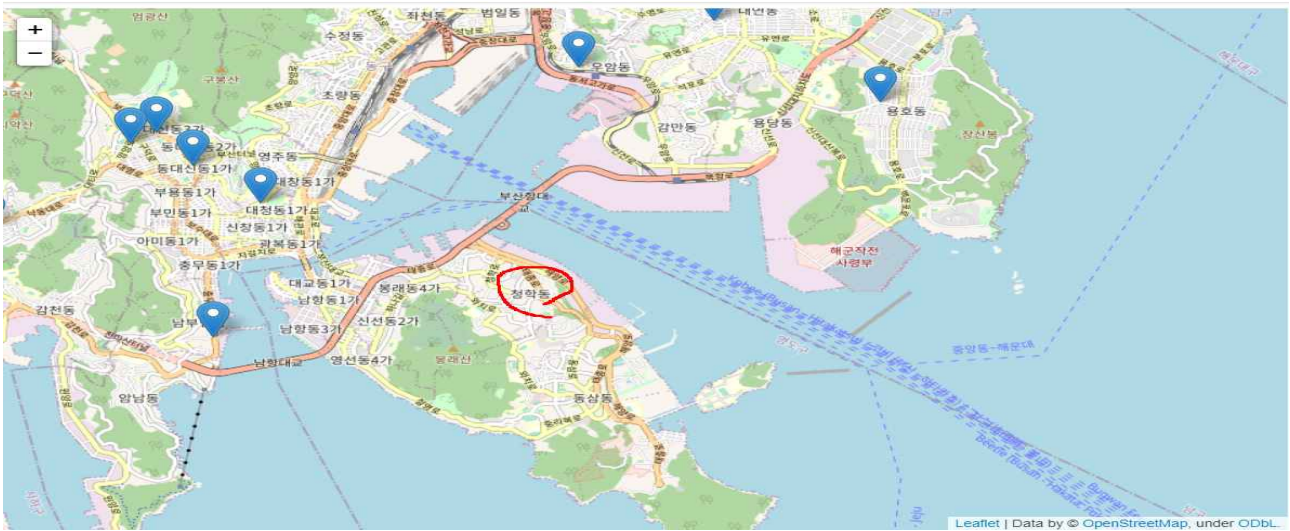
< BBQ치킨 - 부평동(중구) >



```
avg_year = sum(years)/10 # x의 평균
avg_midle_pop = sum(midle_pop)/10 # y의 평균
divisor = sum([(1-avg_year)**2 for i in years])
dividend = top(years, avg_year, midle_pop, avg_midle_pop)
a = dividend / divisor # 최소 제곱해의 기울기
b = avg_midle_pop - (avg_year * a) # 최소 제곱해의 y절편
predict_2020 = a * 2020 + b
if avg_midle_pop < predict_2020:
    print("부평동에서 BBQ 프랜차이즈 신장개업이 가능합니다.")
else:
    print("부평동에서 BBQ 프랜차이즈 신장개업이 불가능합니다.")
```

부평동에서 BBQ 프랜차이즈 신장개업이 불가능합니다.

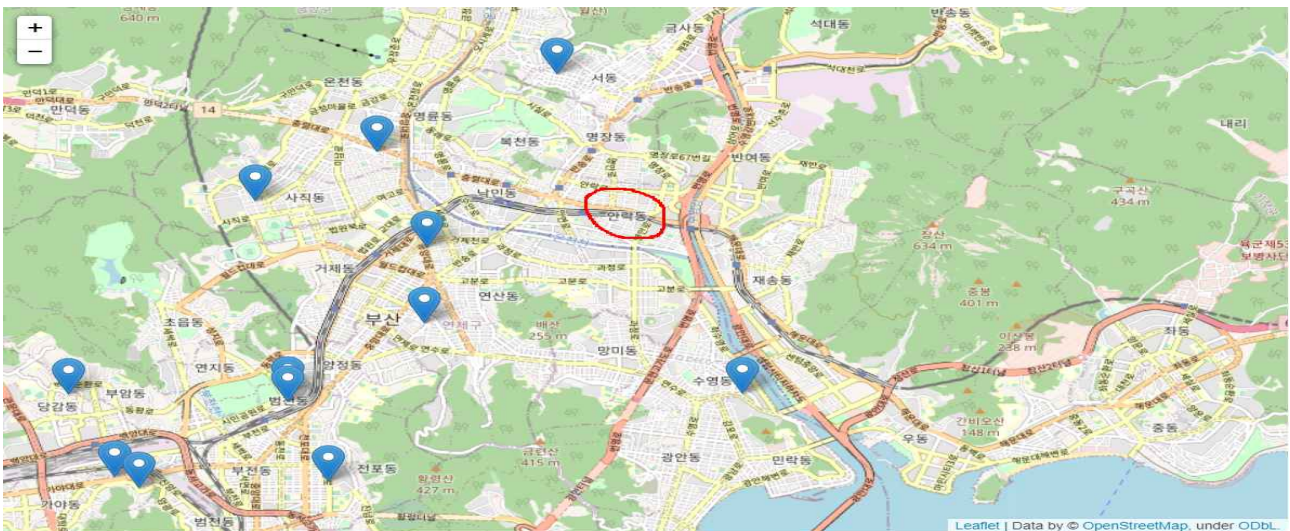
< 멕시코치킨 - 청학동(영도구) >



```
avg_year = sum(years)/10 # x의 평균
avg_young_pop = sum(young_pop)/10 # y의 평균
divisor = sum([(1-avg_year)**2 for i in years])
dividend = top(years, avg_year, young_pop, avg_young_pop)
a = dividend / divisor # 최소 제곱해의 기울기
b = avg_young_pop - (avg_year * a) # 최소 제곱해의 y절편
predict_2020 = a * 2020 + b
if avg_young_pop < predict_2020:
    print("청학동에서 멕시코 프랜차이즈 신장개업이 가능합니다.")
else:
    print("청학동에서 멕시코 프랜차이즈 신장개업이 불가능합니다.")
```

청학동에서 멕시코 프랜차이즈 신장개업이 불가능합니다.

< 네네치킨 - 안락동(동래구) >



```
avg_year = sum(years)/10 # x의 평균
avg_dong_pop = sum(dong_pop)/10 # y의 평균
divisor = sum([(1-avg_year)**2 for i in years])
dividend = top(years, avg_year, dong_pop, avg_dong_pop)
a = dividend / divisor # 최소 제곱해의 기울기
b = avg_dong_pop - (avg_year * a) # 최소 제곱해의 y절편
predict_2020 = a * 2020 + b
if avg_dong_pop < predict_2020:
    print("청학동에서 멕시코 프랜차이즈 신장개업이 가능합니다.")
else:
    print("청학동에서 멕시코 프랜차이즈 신장개업이 불가능합니다.")
```

청학동에서 멕시코 프랜차이즈 신장개업이 불가능합니다.

프로젝트를 하면서 파이썬으로는 정말 다양한 것을 할 수 있다고 느꼈다. 또한 내가 직접 정한 주제를 수행하는 과정에서 막히는 부분을 검색을 통해서 직접 해결해가는 과정이 꽤 재미있었다. 마지막으로 능력 부족으로 이 프로젝트를 원하는 대로 하지 못한 점이 아쉬웠다.