



우리 어디서 만날래?

- 어디서 만날지 고민하는 우리를 위한 프로젝트 -



알잘딱깔센 팀

김채은
배지은
이빈



CONTENTS

1



주제 선정 배경

2



데이터 정의 및 처리

3



분석 결과 및 활용

4



기대효과와 한계점

5



질문 Q&A



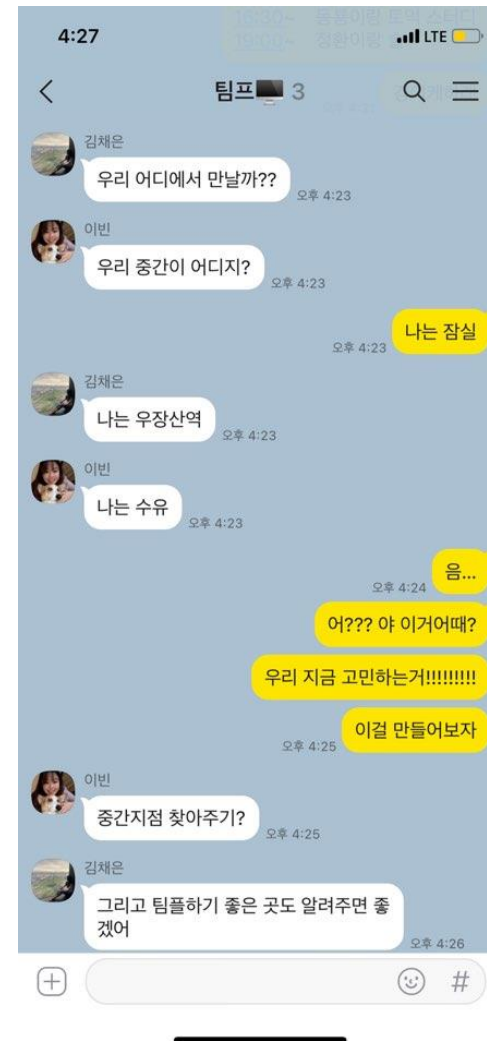
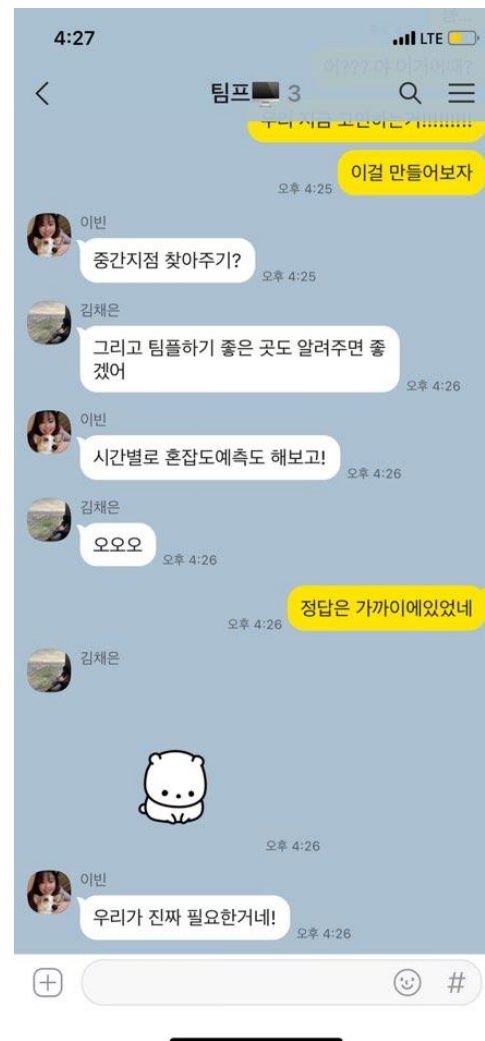


1

주제 선정 배경

1

주제 선정 배경





2

데이터 정의 및 처리

- 지하철 승하차 데이터
- 역별 핫플레이스
- 지하철역 위도, 경도
- 회귀분석 / KNN 알고리즘

2

데이터 정의 및 처리

지하철 역 데이터
핫플 데이터
서울지하철 승하차 데이터

1. 지하철역 주소 데이터 (출처 : 서울교통공사)

- 1 ~9호선 역 중, 서울에 소재하고있는 318개의 역 주소 수집
- 주소를 이용해 위도, 경도로 변환
- 환승역이 존재하는 경우 중복 제거

```
In [3]: sub=pd.read_excel("d:/subway.xlsx",header=0)
...: sub
Out[3]:
```

	호선	역이름	위도	경도
0	1	도봉산	37.689745	127.046153
1	1	도봉	37.679246	127.045478
2	1	방학	37.667678	127.044230
3	1	창동	37.653964	127.047371
4	1	녹천	37.644857	127.051235
...
313	9호선	송파나루	37.510276	127.112224
314	9호선	한성백제	37.515419	127.120637
315	9호선	올림픽공원	37.516466	127.131107
316	9호선	둔촌오륜	37.519296	127.138675
317	9호선	중앙보훈병원	37.530585	127.148043



2

데이터 정의 및 처리

지하철 역 데이터
핫플 데이터
서울지하철 승하차 데이터

2. 지하철 역 별 핫 플레이스 데이터 (직접 수집 - 네이버 지도)

- 테마를 데이트, 팀플, 술로 구분하고 각 테마별로 추천할만한 장소 3곳씩 선정

데이트 : 식당, 카페, 놀거리

팀플 : 프랜차이즈 카페

술약속 : 식당



2

데이터 정의 및 처리

지하철 역 데이터
핫플 데이터
서울지하철 승하차 데이터

3. 서울지하철 승,하차 데이터 (출처 : 서울열린데이터광장)

- 2시간 간격으로 데이터를 합침.
- 하차 인원수 기준으로 혼잡도를 예측함.

여유 : 2000명 이하

보통 : 2001~3000 명

혼잡 : 3001명 이상



Unnamed: 1	역명	승객수	시간대
0	가락시장	488.660944	1
1	가산디지털단지	480.469026	1
2	강남	584.513513	1
3	강남구청	449.139130	1
4	강동	484.930435	1
5	강동구청	455.513043	1
6	강변	523.769231	1
7	개롱	483.596491	1
8	개화산	474.661017	1
9	거여	502.681034	1
10	건대입구	476.214286	1
11	경복궁	527.067797	1
12	경찰병원	509.192983	1
13	고덕	505.672566	1
14	고려대	476.666667	1
15	고속터미널	495.216814	1
16	공덕	483.025751	1
17	공릉	459.601770	1



2

데이터 정의 및 처리

단계별 데이터 처리

1단계: 중간지점 찾기

1-1) 입력 받은 두 개 역의 위도, 경도의 중간 지점을 계산

1-2) 역 입력 받고 해당 위도 경도 찾기

```
place1, place2 = input('★ 출발 역을 입력해주세요 (/으로 구분) ★ ##_##역 이름 1 / 역 이름 2 : ').split('/')
```

```
x1 = data.loc[list(data['subname']).index(place1), 'latitude']  
y1 = data.loc[list(data['subname']).index(place1), 'longitude']  
x2 = data.loc[list(data['subname']).index(place2), 'latitude']  
y2 = data.loc[list(data['subname']).index(place2), 'longitude']
```

★ 출발 역을 입력해주세요 (/으로 구분) ★

역 이름 1 / 역 이름 2 : 고속터미널/쌍문

1-3) 중간지점 찾기

```
def midpoint_euclidean(x1,y1,x2,y2):  
    dist_x = abs(x1-x2) / 2.  
    dist_y = abs(y1-y2) / 2.  
    res_x = x1 - dist_x if x1 > x2 else x2 - dist_x  
    res_y = y1 - dist_y if y1 > y2 else y2 - dist_y  
    return res_x, res_y
```

```
value1, value2 = midpoint_euclidean(x1,y1,x2,y2)
```



2

데이터 정의 및 처리

단계별 데이터 처리

1단계: 중간지점 찾기

1-2) 중간 지점의 위도, 경도를 기준으로 후보 역 뽑아내기

```
# lat: 전체 위도 소수점 3자리  
# latitude_like: value1의 소수점 3자리와 같은 위도 뽑아내기
```

```
lat=[]  
for i in latitude:  
    lat.append(float(str(i)[0:6]))  
  
latitude_like=[]  
for i in lat:  
    if float(str(value1)[0:6]) == i:  
        latitude_like.append(i)
```

```
# 2-2) 위도 기준 추려진 경도 중에 거리 계산해서 가까운 위치 찾기
```

```
# long: 전체 경도 소수점 3자리  
long=[]  
for i in longitude:  
    long.append(float(str(i)[0:7]))
```

```
# long이랑 lat을 data 데이터프레임에 열로 추가  
data['long']=long  
data['lat']=lat
```

```
# latitude_like에 있는 위도를 값으로 갖는 것의 경도 뽑기  
longitude_like=data.loc[data['lat'] == latitude_like[0], 'longitude']
```

```
# 후보군의 위도 값을 뽑기  
lat_cand=[]  
for i in longitude_like:  
    lat_cand.append(data.loc[data['longitude'] == i, 'latitude'])
```

```
# 위도 후보: lat_cand  
# 경도 후보: longitude_like
```



2

데이터 정의 및 처리

단계별 데이터 처리

1단계: 중간지점 찾기

1-3) 유클리디안 거리 계산을 통해 후보 역 중 가장 가까운 역을 출력

#2-4) 최소거리 함수 적용

#후보들의 위도와 경도를 데이터프레임으로 합치기

```
lat_lst=[]
for i in range(len(lat_cand)):
    lat_lst.append(list(lat_cand[i])[0])

cand=DataFrame({'lat':lat_lst, 'long':list(longitude_like)})

# 거리 함수 적용
dis=[]
for i,j in zip(cand['lat'],cand['long']):
    dis.append(GeoUtil.get_euclidean_distance(value2, value1, j, i))

## 가장 작은 값
dis.index(min(dis))

## 가장 가까운 역 출력
mid=data.loc[data['latitude']==cand.iloc[dis.index(min(dis))][0]]['subname']
mid=list(mid)[0]
data.loc[data['longitude']==cand.iloc[dis.index(min(dis))][1]]['subname']
```

79 신설동
Name: subname, dtype: object

2

데이터 정의 및 처리

단계별 데이터 처리

2단계: 중간 지점 역의 핫플 추천하기

만남의 목적을 입력 받아 해당 핫플 추천

★ 만남의 목적 ★

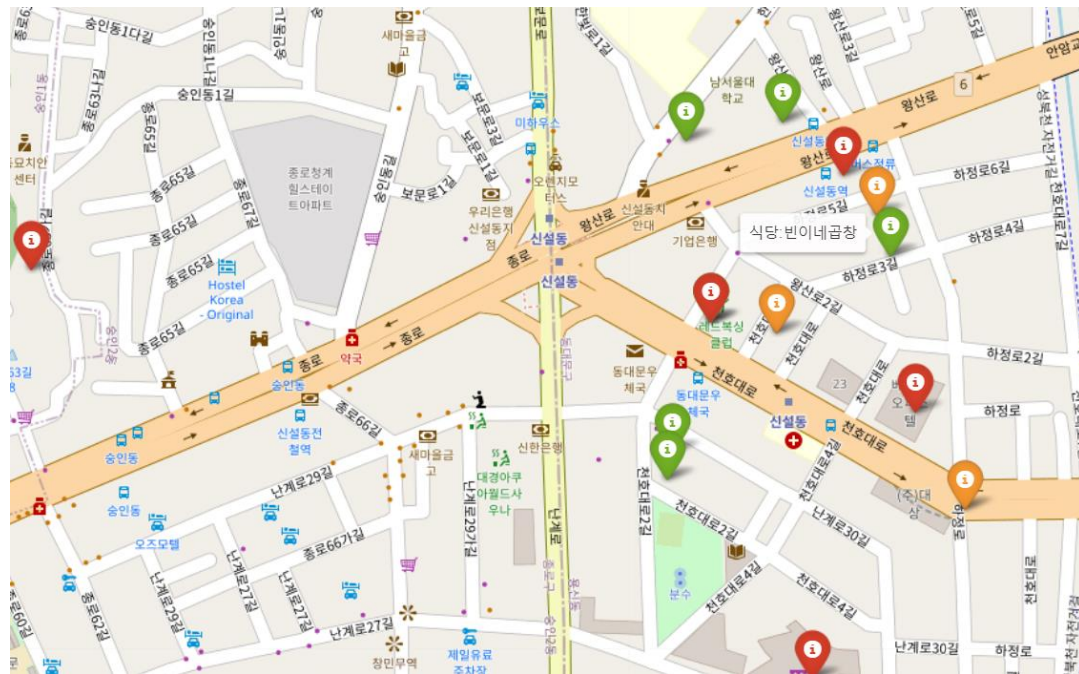
눈누난나 데이트 (데이트): 1

모여라 취해보자 (술모임): 2

마밀라피나타파이 (팀플): 3

숫자를 입력해주세요: 1

	분류	상호명
261	놀거리	청계천박물관
1187	놀거리	한스가죽공방
1188	놀거리	아름다운도보여행
260	놀거리	더 타이 테라피
259	놀거리	서울풍물시장
1189	놀거리	묘각사 템플스테이
255	식당	빈이네곱창
1181	식당	수타생고기
1182	식당	육전식당 1호점
1183	식당	오픈더파스타
254	식당	즐거움맛 돈까스
253	식당	육전식당 2호점
257	카페	캔캔커피
256	카페	플라워카페 하정
1184	카페	플라워카페 하정
1185	카페	캔캔커피
1186	카페	익골
258	카페	익골



2

데이터 정의 및 처리

단계별 데이터 처리

3단계: 승객수 예측하기 - 회귀분석

3-1) 시간대, 지하철역, 평일/주말 별 승객수 데이터를 학습하여 회귀모델 구축

3-2) 범주형 데이터인 시간대, 지하철역, 평일/주말을 더미변수로 변환

3-3) 입력 값에 대한 승객수 예측

```
total.columns=['subname', 'dayweek', 'passenger']
sub_variable = pd.get_dummies(total[['subname', 'dayweek', 'passenger']])
X = sub_variable[sub_variable.columns.difference(['passenger'])]
y = sub_variable['passenger']
model = LinearRegression()
model.fit(X, y)
```

```
# 입력 데이터 받기
colname=list(X.columns)
X_test=[]
for i in range(len(colname)):
    X_test.append(0)
X_test[colname.index('subname_'+mid)]=1
if dayweek_num == 0:
    X_test[colname.index('dayweek_평일')]=1
elif dayweek_num == 1:
    X_test[colname.index('dayweek_주말')]=1
```

```
# 입력 데이터의 승객수 예측값
y_pred = model.predict(np.reshape(X_test,(1,-1)))
```



2

데이터 정의 및 처리

단계별 데이터 처리

4단계: 예측 승객수로 혼잡도 분류 - KNN(K-Nearest Neighbor)

4-1) 시간, 역명을 수치형 데이터로 변환함.

4-2) 하차 승객수 데이터를 이용하여 시간별 혼잡도 예측

```
# 여유/ 보통/ 혼잡 예측
data.loc[data['승객수'] <= 2000, 'con'] = '여유'
data.loc[data['승객수'] > 2000, 'con'] = '보통'
data.loc[data['승객수'] > 3000, 'con'] = '혼잡'
data[data['con'] == '보통']

# 역명을 번호로 바꾸고
A = pd.DataFrame(columns=['역명', 'unique_number'])
A['역명'] = list(data['역명'].unique())
A['unique_number'] = list(range(1, 243))

final = pd.merge(data, A, on='역명')

# knn할꺼만 따로 데이터 프레임 만들고
final1 = final[['승객수', 'unique_number', '시간대', 'con']]

# knn 학습
x_train = np.array(final1.iloc[:, 1:3])
y_train = np.array([1, 2])
label = final1['con']
clf = KNeighborsClassifier(n_neighbors=1)
clf.fit(x_train, label)

# 입력 데이터 만들기: unique_number랑 시간대 입력
X_test_hon = [list(final.loc[final['역명'] == mid, 'unique_number'])

# 입력 데이터의 혼잡도 예측값 출력
result = clf.predict(np.reshape(X_test_hon, (1, 2)))
```

	역명	unique_number
0	가락시장	1
1	가산디지털단지	2
2	강남	3
3	강남구청	4
4	강동	5
5	강동구청	6
6	강변(동서울터미널)	7
7	개롱	8
8	개화산	9
9	거여	10
10	건대입구	11
11	경북궁(정부서울청사)	12
12	경찰병원	13
13	고덕	14
14	고려대(중앙)	15
15	고속터미널	16
16	공덕	17
17	공릉(서울과학기술대)	18
18	광나루(장신대)	19
19	광명사거리	20



3

분석 결과 및 활용

덕선이의 하루 계획에 맞는 중간 지점과
핫플레이스 추천하기

덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

2020년 11월 10일 화요일
덕선이의 계획표

시간	내용
13:00 ~	택이랑 데이트
16:30 ~	동령이랑 토익 스터디
19:00 ~	정환이랑 술

어디서 만나지?
사람들은 얼마나 많으려나?



덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

계획 1 : 데이트

2020년 11월 10일 화요일

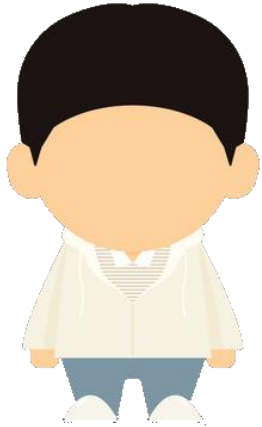
13:00 ~

택이랑 데이트

★ 출발 역을 입력해주세요 (/으로 구분) ★

역 이름 1 / 역 이름 2 : 고속터미널/쌍문

79 신설동
Name: subname, dtype: object



3호선 고속터미널 역 출발

3

1

중간지점
1호선 신설동 역

4



4호선 쌍문 역 출발

우리가 평일에 만날 신설동역의 13~15시간대의 예상 승객수는 4474명이고 예상 혼잡도는 ['여유'] 입니다

덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

계획 1 : 데이트

2020년 11월 10일 화요일



1호선 신설동

★ 만남의 목적 ★

- 눈누난나 데이트 (데이트) : 1
- 모여라 취해보자 (술모임) : 2
- 마밀라피나타파이 (팀플) : 3

숫자를 입력해주세요: 1

- | | 분류 | 상호명 |
|------|-----|-----------|
| 261 | 놀거리 | 청계천박물관 |
| 1187 | 놀거리 | 한스가죽공방 |
| 1188 | 놀거리 | 아름다운도보여행 |
| 260 | 놀거리 | 더 타이 테라피 |
| 259 | 놀거리 | 서울풍물시장 |
| 1189 | 놀거리 | 묘각사 템플스테이 |
| 255 | 식당 | 빈이네곱창 |
| 1181 | 식당 | 수타생고기 |
| 1182 | 식당 | 육전식당 1호점 |
| 1183 | 식당 | 오픈더파스타 |
| 254 | 식당 | 즐거움맛 돈까스 |
| 253 | 식당 | 육전식당 2호점 |
| 257 | 카페 | 캔커피 |
| 256 | 카페 | 플라워카페 하정 |
| 1184 | 카페 | 플라워카페 하정 |
| 1185 | 카페 | 캔커피 |
| 1186 | 카페 | 익골 |
| 258 | 카페 | 익골 |

13:00 ~	택이랑 데이트
---------	---------



덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

계획 2 : 팀플

2020년 11월 10일 화요일

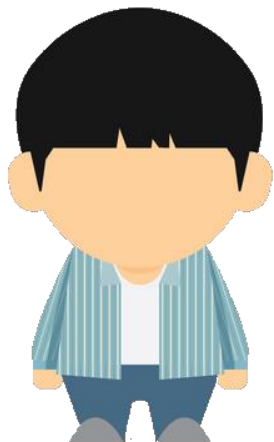
16:30 ~

동릉이랑 토익 스터디

★ 출발 역을 입력해주세요 (/으로 구분) ★

역 이름 1 / 역 이름 2 : 신설동/여의도

210 대흥
Name: subname, dtype: object



5호선 여의도 역 출발

5

6

중간지점

6호선 대흥역

1



1호선 신설동 역 출발

우리가 평일에 만날 대흥역의 15~17시간대의 예상 승객수는 1921명이고 예상 혼잡도는 ['보통'] 입니다

덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

계획 2 : 팀플

2020년 11월 10일 화요일

16:30 ~

동릉이랑 토익 스터디



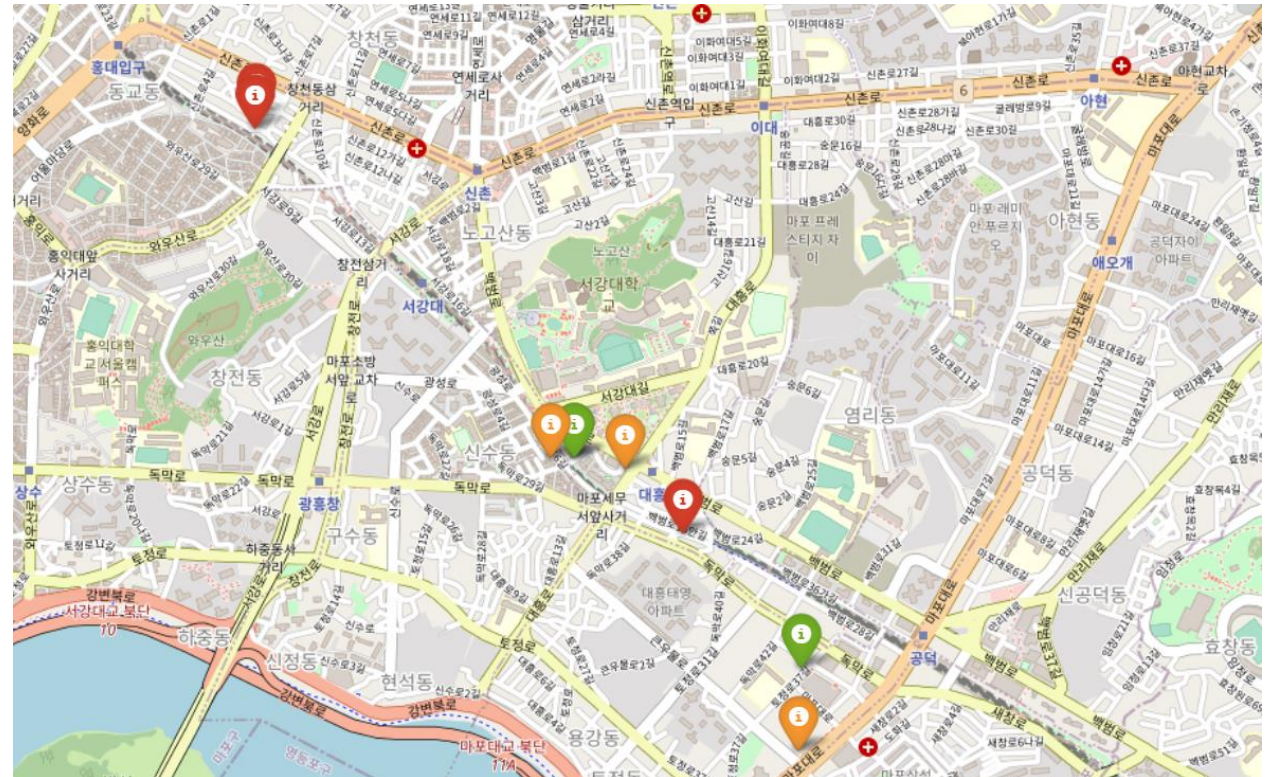
6호선 대흥

★ 만남의 목적 ★

눈누난나 데이트 (데이트): 1
모여라 취해보자 (술모임): 2
마밀라피나타파이 (팀플): 3

숫자를 입력해주세요: 3

	분류	상호명
3158	카페	스타벅스
3159	카페	파리바게뜨
3160	카페	우리동네나무그늘



덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

계획 3 : 술모임

2020년 11월 10일 화요일

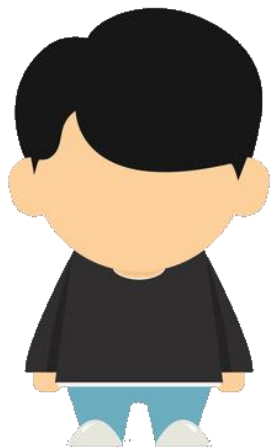
★ 출발 역을 입력해주세요 (/으로 구분) ★

역 이름 1 / 역 이름 2 : 김포공항/대흥

11 서울역
Name: subname, dtype: object

19:00 ~

정환이랑 술



9호선 김포공항 역 출발

9

4

중간지점

4호선 서울역

6



6호선 대흥 역 출발

우리가 평일에 만날 서울역역의 19~21시간대의 예상 승객수는 17270명이고 예상 혼잡도는 ['보통'] 입니다

덕선이의 하루 계획에 맞는 중간 지점과 핫플레이스 추천하기

계획 3 : 술모임

2020년 11월 10일 화요일

19:00 ~

정환이랑 술



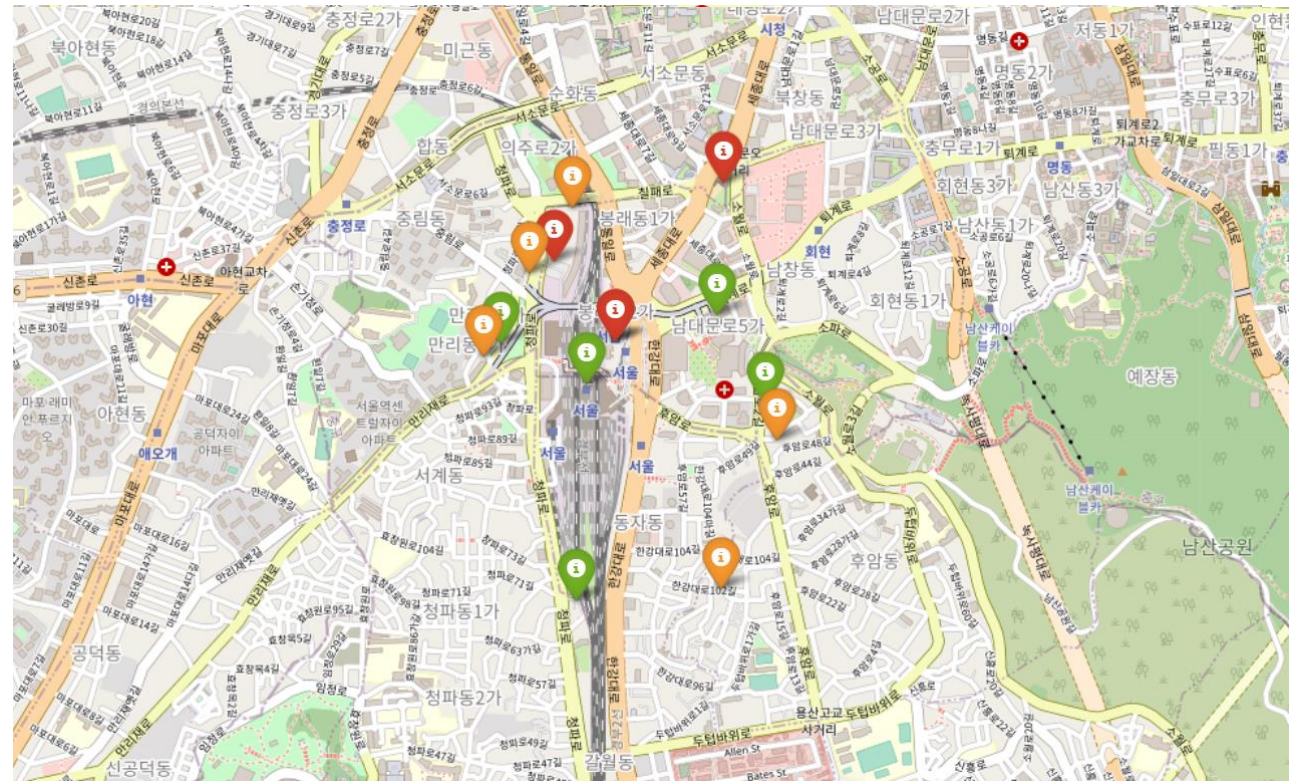
4호선 서울역

★ 만남의 목적 ★

눈누난나 데이트 (데이트): 1
모여라 취해보자 (술모임): 2
마밀라피나타파이 (팀플): 3

숫자를 입력해주세요: 2

	분류	상호명
173	식당	미도리
174	식당	서울리스타
175	식당	MANRI199
2030	식당	도동집
2031	식당	서울리스타
2032	식당	와인살롱





4

한계점



4

한계점

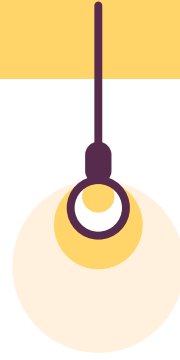
✓ 한계점

2명의 경우, 직선으로 중간지점을 찾을 수 있는데,
3명 이상일 경우 중점을 찾는 방법을 달리 적용해야함으로 추후 보완 예정.

핫플레이스 데이터의 경우 우리가 직접수집
BUT, 지속적인 업데이트를 위해서는 크롤링을 이용하는 것이 좋을 듯.

서울 이외의 지역의 지하철 승/하차 데이터를 찾기 어려웠음.
(용인,분당,인천 등)





“THANK YOU FOR WATCHING”

