

In [1]: `import pandas as pd`

```
grand_1 = pd.read_csv('C:/1_grandata_seoul_single.csv')
grand_2 = pd.read_csv('C://2_grandata_한글.csv', index_col =0)
grand_1 = grand_1.drop(['기준시점'],axis=1)
grand_1 = grand_1.groupby('아파트단지코드').sum()
grand_1 = grand_1.reset_index()
grand_1_2 = pd.merge(grand_1,grand_2)
grand_1_2
```

Out[1]:

	아파트 단지코 드	라 이 프 스 테 이 지	요식이용 금액	유흥이 용금액	백화점이 용금액	대형할인 점이용금 액	소형유통 점이용금 액	의류/잡 화이용금 액	스포츠/ 문화/레 저이용금 액	숙박이 용금액
0	A100064	4	28357000	367000	5115000	10665000	19474000	5128000	5354000	526000
1	A100070	4	38529000	5068000	3967000	10025000	26564000	1590000	10145000	1816000
2	A100079	4	29840000	287000	4257000	10090000	19777000	1626000	4085000	909000
3	A100086	4	74321000	5025000	6308000	8156000	41960000	5354000	17270000	2833000
4	A100163	4	11962000	27000	2519000	2422000	10843000	649000	2597000	267000
...	...	...	...	...	...	...	...	...	...	...
3605	A99883	4	94290000	2704000	12962000	12147000	61765000	10792000	19173000	3183000
3606	A99893	4	41169000	1735000	3242000	5029000	25728000	4589000	12297000	1682000
3607	A99911	4	33623000	429000	10684000	7730000	23120000	4402000	6861000	585000
3608	A99915	4	11926000	594000	2762000	1253000	8923000	318000	2883000	2414000
3609	A99978	4	32380000	259000	4072000	5538000	17540000	3894000	6660000	879000

3610 rows × 116 columns



In [2]: `home = pd.read_excel('C:/fsecdata/수정본 포함(한투, 그랜데이터)/백업/bizcell코드_매칭테이`

In [3]: `grand_1_2 = pd.merge(grand_1_2,home,left_on = '아파트단지코드',right_on = 'HOME_BIZ_CEL  
grand_1_2`

Out[3]:

	아파트 단지코드	라이프 스타일 이치	요식이용 금액	유흥이용 금액	백화점이용 금액	대형할인 점이용금액	소형유통 점이용금액	의류/잡 화이용금액	스포츠/ 문화/레 저이용금액	숙박이용 금액
0	A100064	4	28357000	367000	5115000	10665000	19474000	5128000	5354000	526000
1	A100070	4	38529000	5068000	3967000	10025000	26564000	1590000	10145000	1816000
2	A100079	4	29840000	287000	4257000	10090000	19777000	1626000	4085000	909000
3	A100086	4	74321000	5025000	6308000	8156000	41960000	5354000	17270000	2833000
4	A100163	4	11962000	27000	2519000	2422000	10843000	649000	2597000	267000
...	...	...	...	...	...	...	...	...	...	...
3605	A99883	4	94290000	2704000	12962000	12147000	61765000	10792000	19173000	3183000
3606	A99893	4	41169000	1735000	3242000	5029000	25728000	4589000	12297000	1682000
3607	A99911	4	33623000	429000	10684000	7730000	23120000	4402000	6861000	585000

	아파트 단지코드	라이프 스테이지	요식이용 금액	유흥이용 금액	백화점이용 금액	대형할인 점이용금액	소형유통 점이용금액	의류/잡 화이용금액	스포츠/ 문화/레저이용 금액	숙박이용 금액
3608	A99915	4	11926000	594000	2762000	1253000	8923000	318000	2883000	2414000
3609	A99978	4	32380000	259000	4072000	5538000	17540000	3894000	6660000	879000

3610 rows × 121 columns

```

In [4]: grand_1_2 = grand_1_2.drop(['HOME_BIZ_CELL', '시도명', '아파트단지코드', '행정동코드', '행정동명'])

In [5]: grand_1_2 = grand_1_2.drop(['라이프스테이지'], axis=1)
grand_1_2 = grand_1_2.groupby('시군구명').median()

In [6]: grand_2.columns

Out[6]: Index(['아파트단지코드', '고객수', '하이앤드소비수준', '하이앤드소득수준', '하이앤드법인대표/전문직', '추정연소득',
              '연간카드소비금액', '순자산평가금액', '최근차량할부약정액', '차량보유(국산)', '차량보유(수입)', '20대',
              '30대', '40대', '50대', '60대', '70대이상', '싱글', '신혼및어린이자녀', '청소년자녀', '성인자녀',
              '실버', '명품구매여부', '앱구매등급', '디지털음악이용', '게임이용등급', '골프이용등급', '애완동물관심등급',
              '국내여행등급', '자동차이용여부', '대형할인점', '백화점', '브랜드', '슈퍼마켓', '아울렛', '고궁', '공연장',
              '관광지', '국립공원', '관광_놀이시설', '동물원', '관광_레프팅', '미술관', '박물관/기념관', '식물원',
              '자연휴양림', '골프장', '낚시터', '레저_놀이시설', '레저_레프팅', '볼링장', '스키장', '실내골프', '체험장',
              '캠핑장', '헬스스포츠'],
              dtype='object')

In [7]: grand_1_2= grand_1_2.drop(['행정코드', '고객수', '20대', '30대', '40대', '50대', '60대', '70대', '싱글', '실버'], axis=1)

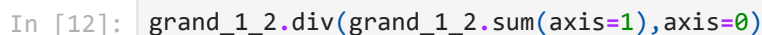
In [8]: from sklearn.metrics.pairwise import cosine_similarity

total_cosine = cosine_similarity(grand_1_2, grand_1_2)

In [9]: ind = grand_1_2.index
cosine_result=pd.DataFrame(total_cosine, columns=ind, index= ind)
```

```
In [11]: import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(20,20))
ax = sns.heatmap(cosine_result, annot=True, fmt='.4')
plt.title('구별 코사인 유사도 전체 column')
plt.show()
```



Out[12]:

	요식이용 금액	유흥이용 금액	백화점이 용금액	대형할인 점이용금 액	소형유통 점이용금 액	의류/잡 화이용금 액	스포츠/ 문화/레 저이용금 액	숙박이용 금액	여행이용 금액	교통이용 금액
시 군 구 명										
강 남 구	0.032482	0.000373	0.008822	0.002563	0.014082	0.005275	0.007466	0.001700	0.000751	0.002412
강 동 구	0.030743	0.000435	0.004835	0.004519	0.016546	0.003881	0.004969	0.000921	0.000368	0.001543
강 북 구	0.027275	0.000513	0.004581	0.004103	0.017190	0.003221	0.004954	0.000985	0.000475	0.000924
강 서 구	0.029273	0.000386	0.005226	0.004737	0.017387	0.002789	0.005035	0.000914	0.000776	0.001506
관 악 구	0.030635	0.000295	0.003847	0.003278	0.019912	0.004144	0.004552	0.000928	0.000511	0.001624
광 진 구	0.032769	0.000271	0.007329	0.003537	0.017194	0.003477	0.005740	0.000942	0.000595	0.001747
구 로 구	0.028459	0.000448	0.005849	0.005317	0.019267	0.003224	0.004679	0.001026	0.000550	0.001662
금 천 구	0.024977	0.000491	0.003431	0.007789	0.017574	0.003880	0.004952	0.000963	0.000262	0.001539
노 원 구	0.024572	0.000412	0.003985	0.003927	0.015513	0.002689	0.004479	0.000826	0.000437	0.001278
도 봉 구	0.031052	0.000560	0.004848	0.006137	0.018701	0.002929	0.005094	0.001087	0.000467	0.001327
동 대 문 구	0.028835	0.000528	0.004405	0.004669	0.020310	0.003444	0.005286	0.001293	0.000503	0.001801
동 작 구	0.029753	0.000402	0.005868	0.003743	0.015299	0.003480	0.005259	0.001244	0.000617	0.002199

	요식이용 금액	유흥이용 금액	백화점이 용금액	대형할인 점이용금 액	소형유통 점이용금 액	의류/잡 화이용금 액	스포츠/ 문화/레 저이용금 액	숙박이용 금액	여행이용 금액	교통이용 금액
시 군 구 명										
마 포 구	0.030327	0.000295	0.005378	0.004580	0.016292	0.003523	0.005997	0.001107	0.000584	0.001964
서 대 문 구	0.026802	0.000339	0.004016	0.004120	0.017981	0.002833	0.004715	0.001045	0.000697	0.001945
서 초 구	0.033139	0.000272	0.010816	0.003950	0.014682	0.004085	0.007276	0.001440	0.000714	0.002034
성 동 구	0.031166	0.000396	0.008055	0.004867	0.016882	0.003879	0.006315	0.001169	0.000743	0.001888
성 북 구	0.027238	0.000343	0.006351	0.004937	0.018074	0.003113	0.004286	0.000995	0.000441	0.001587
송 파 구	0.029795	0.000429	0.007827	0.004115	0.016010	0.003501	0.006572	0.001198	0.000570	0.001932
양 천 구	0.027893	0.000395	0.006618	0.004079	0.018324	0.002803	0.005265	0.000877	0.000665	0.001711
영 등 포 구	0.029689	0.000420	0.007000	0.006247	0.016776	0.003928	0.005308	0.001048	0.000619	0.002181
용 산 구	0.026833	0.000265	0.011837	0.003908	0.014533	0.003939	0.007258	0.001816	0.000638	0.003031
은 평 구	0.027841	0.000362	0.005170	0.003948	0.016360	0.002337	0.004719	0.000771	0.000653	0.001383
종 로 구	0.033761	0.000424	0.010203	0.002920	0.025326	0.003924	0.007853	0.001858	0.000669	0.002657
중 구	0.032287	0.001180	0.007793	0.005030	0.016813	0.004050	0.005607	0.001864	0.001375	0.004188
중 랑 구	0.033502	0.000722	0.004204	0.005993	0.022264	0.004101	0.006009	0.001088	0.000729	0.001622

25 rows × 101 columns

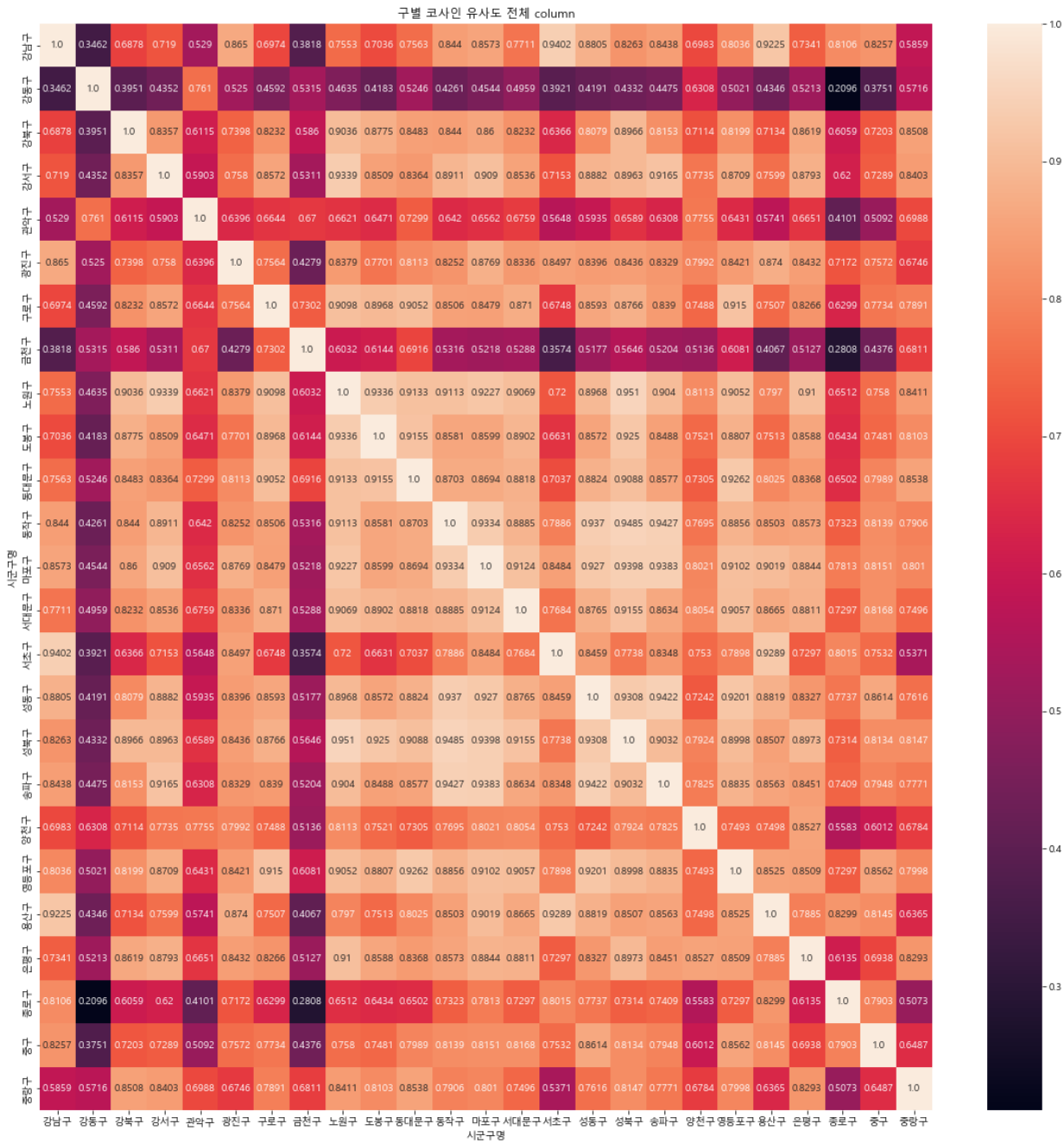
```
In [13]: from sklearn.preprocessing import MinMaxScaler

minmax_scaler = MinMaxScaler()
minmax_scaled = minmax_scaler.fit_transform(grand_1_2)

In [14]: minmax_scale = cosine_similarity(minmax_scaled,minmax_scaled)
ind = grand_1_2.index
cosine_result=pd.DataFrame(minmax_scale,columns=ind, index= ind)

In [60]: a = pd.DataFrame(minmax_scale)

In [15]: plt.figure(figsize=(20,20))
ax = sns.heatmap(cosine_result,annot=True,fmt='.4')
plt.title('구별 코사인 유사도 전체 column')
plt.show()
```



In [16]: cosine\_result



Out[16]:

시 군 구 명	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
시 군 구 명										
강 남 구	1.000000	0.346214	0.687755	0.719036	0.528974	0.865050	0.697364	0.381813	0.755299	0.703584
강 동 구	0.346214	1.000000	0.395074	0.435163	0.761013	0.524993	0.459178	0.531463	0.463490	0.418301
강 북 구	0.687755	0.395074	1.000000	0.835707	0.611474	0.739779	0.823165	0.585956	0.903618	0.877497
강 서 구	0.719036	0.435163	0.835707	1.000000	0.590287	0.758004	0.857233	0.531086	0.933850	0.850926
관 악 구	0.528974	0.761013	0.611474	0.590287	1.000000	0.639649	0.664436	0.670025	0.662149	0.647085
광 진 구	0.865050	0.524993	0.739779	0.758004	0.639649	1.000000	0.756442	0.427886	0.837920	0.770119
구 로 구	0.697364	0.459178	0.823165	0.857233	0.664436	0.756442	1.000000	0.730185	0.909820	0.896828
금 천 구	0.381813	0.531463	0.585956	0.531086	0.670025	0.427886	0.730185	1.000000	0.603242	0.614370
노 원 구	0.755299	0.463490	0.903618	0.933850	0.662149	0.837920	0.909820	0.603242	1.000000	0.933627
도 봉 구	0.703584	0.418301	0.877497	0.850926	0.647085	0.770119	0.896828	0.614370	0.933627	1.000000
동 대 문 구	0.756292	0.524645	0.848291	0.836351	0.729850	0.811336	0.905226	0.691598	0.913311	0.915502
동 작 구	0.843973	0.426058	0.843981	0.891056	0.641982	0.825230	0.850618	0.531550	0.911272	0.858075
마 포 구	0.857284	0.454411	0.860036	0.908951	0.656229	0.876933	0.847858	0.521793	0.922734	0.859936

시 군 구 명	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
시 군 구 명										
서 대 문 구	0.771134	0.495864	0.823225	0.853559	0.675890	0.833555	0.871044	0.528785	0.906898	0.890247
서 초 구	0.940246	0.392098	0.636551	0.715286	0.564804	0.849652	0.674789	0.357415	0.720003	0.663135
성 동 구	0.880450	0.419057	0.807903	0.888206	0.593525	0.839626	0.859316	0.517746	0.896805	0.857205
성 북 구	0.826311	0.433193	0.896579	0.896322	0.658890	0.843554	0.876577	0.564555	0.950972	0.925047
송 파 구	0.843765	0.447497	0.815339	0.916504	0.630832	0.832892	0.838993	0.520407	0.903966	0.848786
양 천 구	0.698301	0.630840	0.711430	0.773483	0.775524	0.799154	0.748807	0.513594	0.811332	0.752125
영 등 포 구	0.803555	0.502088	0.819889	0.870926	0.643120	0.842077	0.915048	0.608091	0.905185	0.880721
용 산 구	0.922490	0.434645	0.713411	0.759874	0.574090	0.874024	0.750659	0.406687	0.797003	0.751325
은 평 구	0.734102	0.521324	0.861937	0.879300	0.665085	0.843179	0.826639	0.512679	0.910025	0.858765
종 로 구	0.810604	0.209636	0.605942	0.619974	0.410117	0.717230	0.629925	0.280775	0.651218	0.643446
중 구	0.825705	0.375073	0.720290	0.728931	0.509231	0.757236	0.773377	0.437623	0.758049	0.748059
중 랑 구	0.585862	0.571603	0.850793	0.840261	0.698784	0.674561	0.789115	0.681081	0.841055	0.810305

25 rows × 25 columns



```
In [24]: cosine_result.columns
```

```
Out[24]: Index(['강남구', '강동구', '강북구', '강서구', '관악구', '광진구', '구로구', '금천구', '노원구', '도봉구',  
              '동대문구', '동작구', '마포구', '서대문구', '서초구', '성동구', '성북구', '송파구',  
              '양천구', '영등포구',  
              '용산구', '은평구', '종로구', '중구', '중랑구'],  
              dtype='object', name='시군구명')
```

```
In [36]: cosine_result
```

Out[36]:

시 군 구 명	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
시 군 구 명										
강 남 구	1.000000	0.346214	0.687755	0.719036	0.528974	0.865050	0.697364	0.381813	0.755299	0.703584
강 동 구	0.346214	1.000000	0.395074	0.435163	0.761013	0.524993	0.459178	0.531463	0.463490	0.418301
강 북 구	0.687755	0.395074	1.000000	0.835707	0.611474	0.739779	0.823165	0.585956	0.903618	0.877497
강 서 구	0.719036	0.435163	0.835707	1.000000	0.590287	0.758004	0.857233	0.531086	0.933850	0.850926
관 악 구	0.528974	0.761013	0.611474	0.590287	1.000000	0.639649	0.664436	0.670025	0.662149	0.647085
광 진 구	0.865050	0.524993	0.739779	0.758004	0.639649	1.000000	0.756442	0.427886	0.837920	0.770119
구 로 구	0.697364	0.459178	0.823165	0.857233	0.664436	0.756442	1.000000	0.730185	0.909820	0.896828
금 천 구	0.381813	0.531463	0.585956	0.531086	0.670025	0.427886	0.730185	1.000000	0.603242	0.614370
노 원 구	0.755299	0.463490	0.903618	0.933850	0.662149	0.837920	0.909820	0.603242	1.000000	0.933627
도 봉 구	0.703584	0.418301	0.877497	0.850926	0.647085	0.770119	0.896828	0.614370	0.933627	1.000000
동 대 문 구	0.756292	0.524645	0.848291	0.836351	0.729850	0.811336	0.905226	0.691598	0.913311	0.915502
동 작 구	0.843973	0.426058	0.843981	0.891056	0.641982	0.825230	0.850618	0.531550	0.911272	0.858075
마 포 구	0.857284	0.454411	0.860036	0.908951	0.656229	0.876933	0.847858	0.521793	0.922734	0.859936

시 군 구 명	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구
시 군 구 명										
서 대 문 구	0.771134	0.495864	0.823225	0.853559	0.675890	0.833555	0.871044	0.528785	0.906898	0.890247
서 초 구	0.940246	0.392098	0.636551	0.715286	0.564804	0.849652	0.674789	0.357415	0.720003	0.663135
성 동 구	0.880450	0.419057	0.807903	0.888206	0.593525	0.839626	0.859316	0.517746	0.896805	0.857205
성 북 구	0.826311	0.433193	0.896579	0.896322	0.658890	0.843554	0.876577	0.564555	0.950972	0.925047
송 파 구	0.843765	0.447497	0.815339	0.916504	0.630832	0.832892	0.838993	0.520407	0.903966	0.848786
양 천 구	0.698301	0.630840	0.711430	0.773483	0.775524	0.799154	0.748807	0.513594	0.811332	0.752125
영 등 포 구	0.803555	0.502088	0.819889	0.870926	0.643120	0.842077	0.915048	0.608091	0.905185	0.880721
용 산 구	0.922490	0.434645	0.713411	0.759874	0.574090	0.874024	0.750659	0.406687	0.797003	0.751325
은 평 구	0.734102	0.521324	0.861937	0.879300	0.665085	0.843179	0.826639	0.512679	0.910025	0.858765
종 로 구	0.810604	0.209636	0.605942	0.619974	0.410117	0.717230	0.629925	0.280775	0.651218	0.643446
중 구	0.825705	0.375073	0.720290	0.728931	0.509231	0.757236	0.773377	0.437623	0.758049	0.748059
중 랑 구	0.585862	0.571603	0.850793	0.840261	0.698784	0.674561	0.789115	0.681081	0.841055	0.810305

25 rows × 25 columns

In [56]: cosine\_result

Out[56]: Index(['index', '시군구명', '강남구', '강동구', '강북구', '강서구', '관악구', '광진구', '구로구', '금천구', '노원구', '도봉구', '동대문구', '동작구', '마포구', '서대문구', '서초구', '성동구', '성북구', '송파구', '양천구', '영등포구', '용산구', '은평구', '종로구', '중구', '중랑구'], dtype='object', name='시군구명')

In [63]: a.index= ['강남구', '강동구', '강북구', '강서구', '관악구', '광진구', '구로구', '금천구', '동대문구', '동작구', '마포구', '서대문구', '서초구', '성동구', '성북구', '송파구', '용산구', '은평구', '종로구', '중구', '중랑구']

In [65]: a.columns= ['강남구', '강동구', '강북구', '강서구', '관악구', '광진구', '구로구', '금천구', '동대문구', '동작구', '마포구', '서대문구', '서초구', '성동구', '성북구', '송파구', '용산구', '은평구', '종로구', '중구', '중랑구']

In [68]: a\_c =pd.melt(a.reset\_index(),id\_vars='index')

In [85]: a\_c = a\_c.rename(columns={'index':'구\_1','variable':'구\_2'})

In [86]: a\_c.to\_csv('구별\_코사인유사도\_목록.csv')

In [17]: import numpy as np  
import geopandas as gpd

```
sgg_file = 'C:/Users/fsecuser213/Downloads/지리정보/SIG_20220324/sig.shp'
sgg = gpd.read_file(sgg_file,encoding='euc-kr')
seoul = sgg[sgg['SIG_CD'].str.startswith('11')]
seoul_viz = pd.merge(seoul,cosine_result,left_on = 'SIG_KOR_NM',right_on='시군구명')
```

In [22]: import shapely  
seoul\_viz['coords']=seoul\_viz['geometry'].apply(lambda x: x.representative\_point().coords)  
seoul\_viz['coords'] = [coords[0] for coords in seoul\_viz['coords']]

In [23]: for i in seoul\_viz.columns[4:]:  
 ax =seoul\_viz.plot(column=i,legend=True,figsize=(10,8),edgecolor='w',cmap='OrRd')  
 for idx,row in seoul\_viz.iterrows():  
 plt.annotate(text=row['SIG\_KOR\_NM'],xy=row['coords'],horizontalalignment='center')  
 ax.axis('off')  
 ax.set\_title(i)  
 ax.set\_axis\_off()  
 plt.tight\_layout()  
 plt.show()

