

# 1. df\_starbucks (MAIN)

구분: 자체 데이터 (크롤링)

테이블명: df\_starbucks

데이터 출처: [https://www.starbucks.co.kr/store/store\\_map.do](https://www.starbucks.co.kr/store/store_map.do) 크롤링(selenium + bs4)

조인 키: 위도, 경도, 시군구명, 행정동명

테이블 상세:

No.	Column	Count	Dtype	가공 여부 (설명)
0	상호명	675	object	N
1	위도	675	float64	N
2	경도	675	float64	N
3	행정동명	675	object	Y (공공데이터 좌표 기반 반환값)
4	시군구명	675	object	Y (공공데이터 좌표 기반 반환값)
5	도로명주소	675	object	N
6	지번주소	675	object	Y (공공데이터 좌표 기반 반환값)
7	스타벅스	675	int64	Y (스타벅스:1, 그외:0)
8	상권업종대분류명	675	object	Y (df_seoul에서 상권업종소분류명 '카페' 기준 대분류명)
9	상권업종중분류명	675	object	Y (df_seoul에서 상권업종소분류명 '카페' 기준 중분류명)
10	상권업종소분류명	675	object	Y (df_seoul에서 상권업종소분류명 '카페' 기준 소분류명)
11	상가업소번호	675	object	Y (sb + 인덱스 번호)
12	행정동코드	675	object	Y (공공데이터 좌표 기반 반환값)

## 2. df\_seoul (MAIN)

구분: 로우 데이터 (공공데이터 포털)

테이블명: df\_seoul

데이터 출처: <https://www.data.go.kr/data/15083033/fileData.do>

조인 키: 행정동코드(df\_population)

상가업소번호(self-join)

위도, 경도(df\_subway, df\_building\_price)

테이블 상세:

RangeIndex: 540517

No.	Column	Count	Dtype	가공 여부 (설명)
0	상가업소번호	540517 non-null	object	N
1	상호명	540517 non-null	object	N
2	지점명	48544 non-null	object	N
3	상권업종대분류코드	540517 non-null	object	N
4	상권업종대분류명	540517 non-null	object	N
5	상권업종중분류코드	540517 non-null	object	N
6	상권업종중분류명	540517 non-null	object	N
7	상권업종소분류코드	540517 non-null	object	N
8	상권업종소분류명	540517 non-null	object	N
9	표준산업분류코드	540317 non-null	object	N
10	표준산업분류명	540317 non-null	object	N
11	시도코드	540517 non-null	int64	N
12	시도명	540517 non-null	object	N
13	시군구코드	540517 non-null	int64	N
14	시군구명	540517 non-null	object	N
15	행정동코드	540517 non-null	int64	N
16	행정동명	540517 non-null	object	N
17	법정동코드	540517 non-null	int64	N
18	법정동명	540517 non-null	object	N
19	지번코드	540517 non-null	int64	N
20	대지구분코드	540517 non-null	int64	N
21	대지구분명	540517 non-null	object	N
22	지번본번지	540460 non-null	float64	N
23	지번부번지	448378 non-null	float64	N
24	지번주소	540517 non-null	object	N
25	도로명코드	540517 non-null	int64	N
26	도로명	540517 non-null	object	N
27	건물본번지	540516 non-null	float64	N
28	건물부번지	70031 non-null	float64	N
29	건물관리번호	540516 non-null	object	N
30	건물명	265014 non-null	object	N
31	도로명주소	540517 non-null	object	N
32	구우편번호	540517 non-null	int64	N
33	신우편번호	540517 non-null	int64	N
34	동정보	0 non-null	float64	N

35	층정보	370779 non-null	object	N
36	호정보	0 non-null	float64	N
37	경도	540517 non-null	float64	N
38	위도	540517 non-null	float64	N

### 3. df\_population

구분: 가공 데이터 (공공데이터 포털)

테이블명: df\_population

데이터 출처:

- (1) 행정동 인구: <https://www.data.go.kr/data/15097972/fileData.do>

- (2) 행정동 위경도: <https://skyseven73.tistory.com/23>, 누락 데이터는 수기 검색 후 보강.

조인 키: 행정기관코드

테이블 상세:

No.	Column	Count	Dtype	가공 여부 (설명)
0	행정기관코드	426	int64	N
1	기준연월	426	object	N
2	시도명	426	object	N
3	시군구명	426	object	N
4	읍면동명	426	object	N
5	계	426	int64	N
6	0세_6세	426	int64	Y (0세_6세 인구 수 합계)
7	7세_19세	426	int64	Y (7세_19세 인구 수 합계)
8	20세_25세	426	int64	Y (20세_25세 인구 수 합계)
9	26세_30세	426	int64	Y (26세_30세 인구 수 합계)
10	31세_40세	426	int64	Y (31세_40세 인구 수 합계)
11	41세_50세	426	int64	Y (41세_50세 인구 수 합계)
12	51세_60세	426	int64	Y (51세_60세 인구 수 합계)
13	61세_109세	426	int64	Y (61세_109세 인구 수 합계)

## 4. df\_building\_price

구분: 가공 데이터 (공공데이터 포털)

테이블명: df\_building\_price

데이터 출처:

<https://rt.molit.go.kr/pt/xls/xls.do?mobileAt=>

조인 키: 주소, 도로명, 시군구, 지번, 위도, 경도

테이블 상세:

No.	Column	Count	Dtype	가공 여부 (설명)
0	시군구	9906	object	N
1	지번	9906	object	N
2	도로명	9906	object	N
3	용도지역	9906	object	N
4	건축물주용도	9906	object	N
5	전용/연면적(m²)	9906	float64	N
6	거래금액(만원)	9906	object	Y (전용연면적 / 3.3058)
7	주소	9906	object	N
8	위도	9906	float64	N
9	경도	9906	float64	N

## 5. df\_subway

구분: 로우 데이터 (공공데이터 포털)

테이블명: df\_subway

데이터 출처:

<https://www.data.go.kr/data/15099316/fileData.do?recommendDataYn=Y>

조인 키: 역번호, 역명, 위도, 경도

테이블 상세:

No.	Column	Count	Dtype	가공 여부 (설명)
0	연번	289	int64	N
1	호선	289	int64	N
2	역번호	289	int64	N
3	역명	289	object	N
4	위도	289	float64	N
5	경도	289	float64	N

## 6. df\_subway\_on

구분: 가공 데이터 (공공데이터 포털)

테이블명: df\_subway\_on

데이터 출처:

<https://www.data.go.kr/data/15104835/fileData.do>

<https://www.data.go.kr/data/15062858/fileData.do>

조인 키: 역명

테이블 상세:

No.	Column	Count	Dtype	가공 여부 (설명)
0	역명	286	object	N
1	호선명	286	object	N
2	승차_이용객수	286	float64	Y (구분=='승차' 기준 합계)
3	승차_출근시간대	286	float64	Y (구분=='승차' , 07-10 시간대 합계)
4	승차_퇴근시간대	286	float64	Y (구분=='승차' , 17-20 시간대 합계)

## 7. df\_subway\_off

구분: 가공 데이터 (공공데이터 포털)

테이블명: df\_subway\_off

데이터 출처:

<https://www.data.go.kr/data/15104835/fileData.do>

<https://www.data.go.kr/data/15062858/fileData.do>

조인 키: 역명

테이블 상세:

No.	Column	Count	Dtype	가공 여부 (설명)
0	역명	286	object	N
1	호선명	286	object	N
2	승차_이용객수	286	float64	Y (구분=='승차' 기준 합계)
3	승차_출근시간대	286	float64	Y (구분=='승차' , 07-10 시간대 합계)
4	승차_퇴근시간대	286	float64	Y (구분=='승차' , 17-20 시간대 합계)



