

주제 : 소상공인 상권 데이터를 활용한 창업 유망 지역 추천 시스템

활용 데이터 : <https://www.data.go.kr/data/15083033/fileData.do> 최신 소상공인시장진흥공단\_상가(상권)정보

## 1. 문제 정의

프로젝트 주제:

소상공인 상가(상권) 데이터를 분석하여 창업에 유리한 지역을 추천하는 시스템 개발.

목표:

상권 데이터의 업종 분포와 상가 밀집도를 기반으로 창업 유망 지역을 도출.  
경쟁 강도와 상권의 입지 특성을 분석하여 창업 가능성이 높은 위치를 추천.  
창업자들이 직관적으로 이해할 수 있도록 시각화된 분석 결과를 제공.

## 2. 시스템 아키텍처

### 1. 데이터 수집 :

python으로 <https://www.data.go.kr/data/15083033/fileData.do> 최신 소상공인시장진흥공단\_상가(상권)정보 csv파일 형식 불러오기

### 2. 데이터 저장 :

HDFS: 대용량 데이터를 저장하여 분산 처리.

Hive: 정형 데이터를 SQL 형식으로 질의.

### 3. 데이터 처리 :

spark : 분산 데이터 처리하고 상권 특성 분석

### 4. 시각화 :

Matplotlib : 경쟁도 및 밀집도 시각화

Folium : 지도 기반 밀집도 및 추천 지역 표시

## 3. 데이터 수집 방법

소상공인 상권 데이터:

데이터 형식: CSV

주요 컬럼: 상권업종대분류명, 상권업종중분류명, 경도, 위도, 도로명주소, 시도명, 시군구명.

## 4. 데이터 분석 방법

1. 데이터 전처리

- 결측치 처리, 중복 데이터 제거, 업종 분류

2. 상권 경쟁도 분석

- 지역별 동일 업종의 상가 수 계산

3. 상가 밀집도 분석

- 위도, 경도 데이터를 활용한 밀집도 시각화