

스타벅스 vs. 이디야

데이터로 보는 거리와 분포

전체 흐름

데이터 수집

Selenium과 BeautifulSoup을 활용해서 웹 스크래핑

주소 처리

네이버 API 및 Google Maps API를 이용한 주소 변환

DB 저장

MySQL을 이용한 데이터 저장 및 조회

데이터 분석 및 시각화

Folium 지도, 물리적 거리 분석, 상관관계 분석

1. 주요 클래스 및 역할

DBManager

- **역할:** MySQL 데이터베이스 연결, 데이터 저장, 조회, 실행
- **핵심 메서드:**
 - `insert()` → 매장 데이터를 DB에 추가
 - `select_all()` → 특정 테이블의 모든 데이터 조회
 - `execute()` → SQL 문 직접 실행

AddressManager

- **역할:** 매장의 주소 데이터를 다양한 방법으로 변환
- **핵심 메서드:**
 - `isDoro()` → 입력된 주소가 도로명 주소인지 판별
 - `extractDoro()` → 주소에서 도로명만 추출
 - `getDoro()` → 여러 API(Selenium, Google Maps, 네이버)를 활용해 도로명 주소 변환

APIManager

- **역할:** 네이버 API를 이용한 검색 URL 생성 및 데이터 요청
- **핵심 메서드:**
 - `genSearchUrl()` → 네이버 API 검색 URL 생성
 - `getResultOnpage()` → 네이버 API 요청 및 응답 파싱

ScrappingManager

- **역할:** Selenium을 이용해 웹사이트에서 매장 정보를 자동으로 가져옴
- **핵심 메서드:**
 - `openPage()` → 지정된 URL을 Selenium으로 열기
 - `getSoup()` → HTML을 BeautifulSoup 객체로 변환

2. 데이터 수집 과정

Selenium을 이용해 웹사이트에 접속 & 지역 검색



BeautifulSoup으로 매장 이름, 주소 가져오기



Google Maps API로 위도, 경도 가져오기



주소 정제

- isSeoul() : 서울 매장만 필터링
- isDoro(): 도로명 주소인지 확인
- getDoro(): 지번주소를 도로명주소로

2. 데이터 수집 과정

내방역점

서울 서초구 방배동 898-1



교대역점

서울 서초구 서초동 1669-15

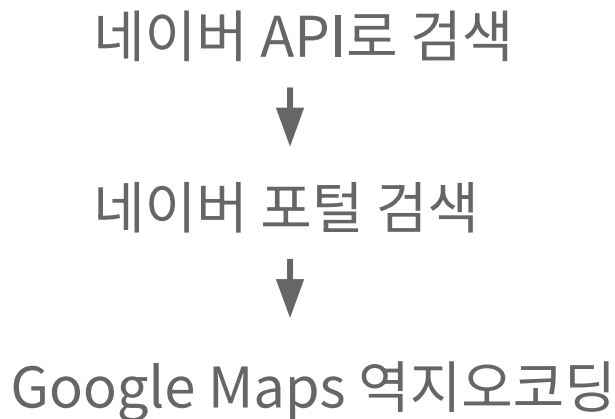


이디야 멤버스에서는 매장 주소를 지번 주소로 제공하는 매장이 많기 때문에,

도로명 주소를 통한 분석을 위해서는 도로명 주소로 변환해 주어야 함

2. 데이터 수집 과정

어떻게 변환할 것인가?



2. 데이터 수집 과정

```
def getDoro(self, address, name=None, lat=None, lng=None):

    if self.isDoro(address):
        return address # 이미 도로명 주소라면 그대로 반환

    # 네이버 API로 검색
    if name:
        doro = self.getDoro_byapi(name)
        if doro:
            return doro

    # 네이버 포털 검색 (Selenium)
    doro = self.getDoro_byportal(address)
    if doro:
        return doro

    # Google Maps API 역지오코딩 (위도, 경도 필요)
    if lat is not None and lng is not None:
        doro = self.getDoro_byRGC(lat, lng)
        if doro:
            return doro

    # 모든 방법을 시도했으나 도로명 주소를 찾지 못한 경우
    return ""
```

3. 데이터 저장

```
mysql> desc ediya;
+-----+-----+-----+-----+-----+-----+
| Field      | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| name       | varchar(32)   | NO   | PRI | NULL    |       |
| address    | varchar(64)   | NO   |     | NULL    |       |
| latitude   | double        | YES  |     | NULL    |       |
| longitude  | double        | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.00 sec)
```

name을 PRIMARY KEY로 지정
-> 중복 데이터 INSERT 방지

4. 데이터 분석 및 시각화

어떻게 분석하였는가?

Folium을 통한 지도 시각화

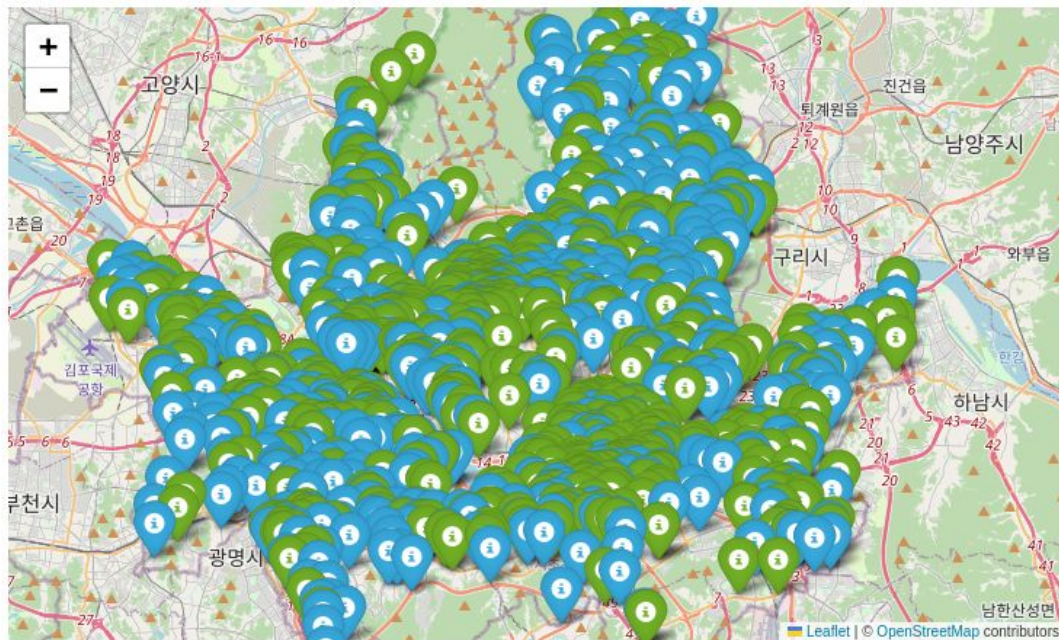
Haversine 공식, KDTree를 통한
물리적 거리 계산

도로명 주소 분석을 통한
상관계수 분석

4. 데이터 분석 및 시각화

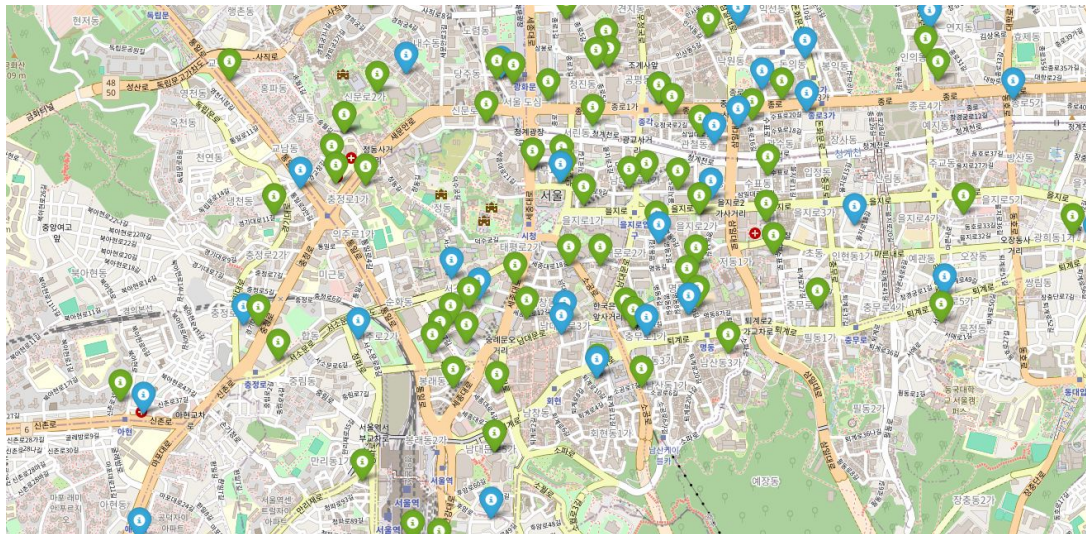
Folium 지도

의미 있는 특징을 찾기 힘들다



4. 데이터 분석 및 시각화

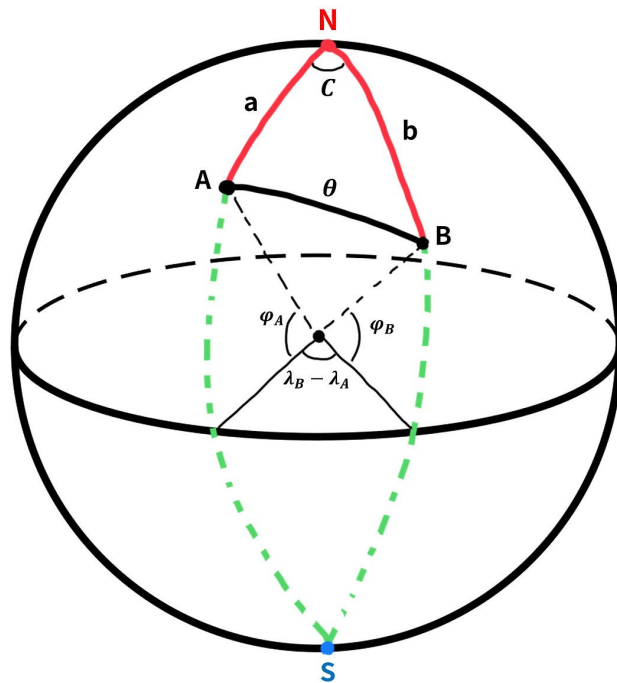
Folium 지도



4. 데이터 분석 및 시각화

물리적 거리 계산
: Haversine

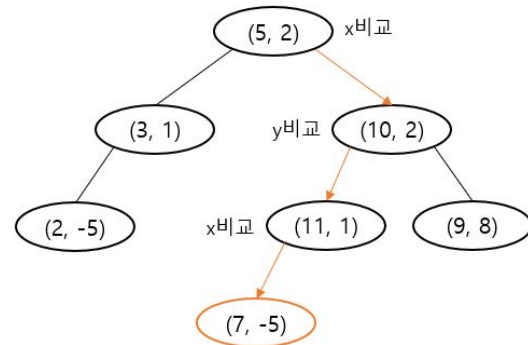
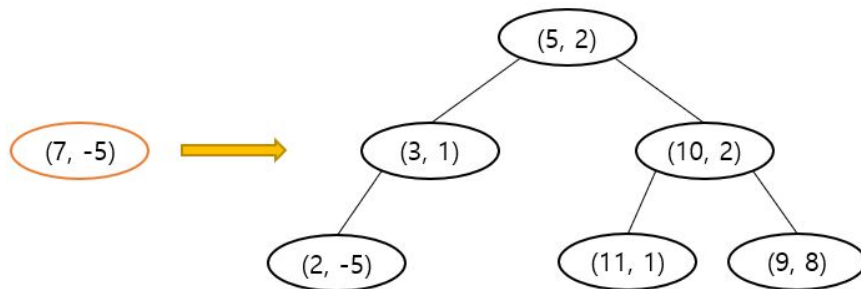
100m 이내	있는 이디야 매장 비율	: 15.32%
300m 이내	있는 이디야 매장 비율	: 47.94%
500m 이내	있는 이디야 매장 비율	: 68.37%



4. 데이터 분석 및 시각화

물리적 거리 계산 : KDTree

100m 이내에 있는 이디야 매장 개수: 65 < 509 (12.77%)
300m 이내에 있는 이디야 매장 개수: 218 < 509 (42.83%)
500m 이내에 있는 이디야 매장 개수: 323 < 509 (63.46%)

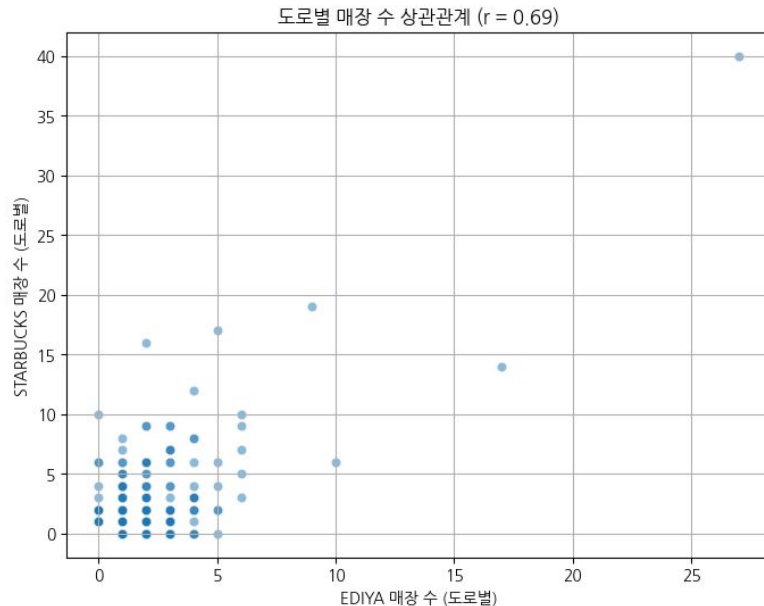


4. 데이터 분석 및 시각화

도로명 주소 분석

상관계수 크기	일반적인 해석
$\pm 0.8 \sim 1.0$	매우 강한 상관관계
$\pm 0.6 \sim 0.8$	강한 상관관계
$\pm 0.4 \sim 0.6$	중간 정도 상관관계
$\pm 0.2 \sim 0.4$	약한 상관관계
$\pm 0.0 \sim 0.2$	매우 약한 상관관계

$$\rho = 0.6936721165295652$$



5. 결론

거리 기반 분석 결과

Haversine 공식에 의한 거리계산과 **KDTree**를 이용한 거리계산의 결과가 유사한 것으로 보아 계산에는 큰 오차가 없는 듯 보임.

- **100m 이내** : 12~15% 정도의 이디야 매장만 스타벅스와 가까이 위치
- **300m 이내**: 42~47% 정도의 이디야 매장이 스타벅스 300m 반경 내에 있음
- **500m 이내**: 63~68%의 이디야 매장이 스타벅스 반경 500m 내에 위치

극단적인 근접 전략보다는 **어느 정도 분리된 입점 전략**을 취하는 듯 보임

약 65% 정도의 이디야 매장이 스타벅스와 같은 상권을 공유한다고 볼 수 있음.

5. 결론

도로명 기반 분석 결과

도로별 이디야 매장과 스타벅스 매장의 상관계수 분석 결과,
0.6937 정도의 상관계수를 보여주었음.

- 0.6937의 상관계수는 **강한 상관관계**를 의미함
(상관관계가 있다고 통계적으로 지지받을 수 있는 수준)
- 즉, 이디야 매장이 많은 도로에서는 스타벅스 매장도 많을 가능성이 높음.

5. 결론

이디야 매장과 스타벅스 매장은 같은 상권을 공유하는 경향을 보이며, 이는 두 브랜드가 유동 인구가 많은 지역을 공략하는 전략을 취하고 있다고 볼 수 있음. 특히, 매장들이 300~500m 이내에 있는 경우가 많아, 같은 상권/생활권 내에서 경쟁할 가능성이 높음.

그러나 다른 요인(유동 인구, 오피스/주거지 밀집도, 역세권 여부 등)도 입점 전략에 영향이 미칠 가능성이 크므로, 직접적인 인과관계를 단정하기는 어려우며 추가적인 분석이 반드시 필요함.