## 相关数据

本系统主要用到的数据是北京市全市的POI数据，以下是数据相关说明。

#### 1. 数据集概述

本系统中使用的北京市POI（Point of Interest）数据集包含北京市范围内的兴趣点信息。该数据为旅游路线规划提供了地理参考和兴趣点推荐，帮助系统更好地为用户提供包括但不限于购物中心、餐饮、酒店、商场、公园等地点的详细信息。

#### 2. 数据来源

北京市POI数据主要来自高德地图，通过调用高德地图API以获取最新的兴趣点数据。为了保证数据的时效性，建议定期更新数据集。

#### 3. 数据结构

POI数据集以CSV格式存储，主要包含以下字段：

* **名称（name）**： 地点的名称
* **大类（large\_category）**：地点分类中的大类别，粒度较粗，例如“风景名胜”、“住宿服务”、“体育休闲服务”、“餐饮服务”等
* **中类（middle\_category）**：地点分类中的中类别，例如“公园广场”、“博物馆”、“影剧院”、“运动场馆”等
* **小类（small\_category）**：地点分类中的小类别，粒度较细，例如“动物园”、“教堂”、“酒吧”、“棋牌室”等
* **地址（address）**： 地点的详细地址
* **省（province）**： 地点所在省份信息
* **市（city）**： 地点所在市信息
* **区（district）**： 地点所在区信息
* **WGS84\_经度（longitude）**：地点在WGS84坐标系下的经度信息
* **WGS84\_纬度（latitude）**：地点在WGS84坐标系下的纬度信息
* **热门度（popularity）**：地点的热门度，基于用户的访问频率计算出的地点受欢迎程度指标，数值在0~1之间，保留9位小数

#### 4. 数据处理流程

1. **数据收集**：通过高德地图API接口获取北京市的POI数据。
2. **数据清理**：清除重复数据、无效数据，并处理缺失值。
3. **数据分类**：根据不同类别进行地点分类，并最终选出景点、酒店、餐厅、娱乐场所、购物五个类别的POI数据。
4. **计算热门度**：结合FourSquare签到数据集统计每个地点的访问频率，并归一化成该地点的热门度分数。
5. **数据存储**：处理后的POI数据存储于数据库中，支持后续的查询和推荐操作。

#### 5. 数据集示例

