

**VAMAP**

OSS 프로젝트 프로젝트 제안서

오픈소스 프로젝트

김효수 교수님

이상민 | 20164204

주석천 | 20151363

이현재 | 20133096

민성재 | 20145223

정유석 | 20141359

# 프로젝트 주제 소개

국가적 정책의 일환으로 수 많은 공공데이터가 생겨나고 있다. 또한 지리적 위치와 연결 되어있는 데이터는 누구나 가지고 있고 누구나 만들 수 있다. 하지만 이런 데이터가 많은 것과는 별개로 분석하는것과 시각화 하는 것은 별개의 문제이다.

기존의 프로그램들은 지도상에서 단순 특정데이터의 밀집도 혹은 분포도만 보여주는 경우가 많다. **하지만, 정작 밀집도를 통해 어떠한 분석결과를 한눈에 알아보기는 쉽지 않다. 따라서, 우리의 개발목표는 사용자가 원하는 위치기반 (위도 및 경도) 데이터를 입력하면 도로, 건물, 지하철역 등의 데이터와 함께 분석하여 각 위치기반 데이터 객체마다의 상대적 가치를 계산하고 시각화 해주는 것을 목표로 한다.** 이때 서울의 길, 건물, 역을 기반으로 개발을 진행할 예정이다

**물론, 특정 상점의 위치기반 데이터가 이 프로그램을 통해 얻어지는 결과 상권분석이라고 판단이 될 수도 있지만, 이 프로그램의 궁극적인 목표는 공공시설이나 자재 등의 위치를 통해 각 시설 및 자재 등이 얼마나 낭비되고 있는지 혹은 필요하다고 예상되는 곳에 없는지 등의 정보를 제공하는 것이다.**

# 역할분담

**이현재[팀장]**

* Readme, Report 작성, 발표준비 및 발표 등의 문서작업
* 팀원별로 작업지시 및 업무 세부화
* 작업진행이 부족한 부분에 보조 개발자로 투입

**주석천**

* 데이터별 예상되는 시나리오 구상 및 설계
* 모든 데이터 전처리 및 가공
* 응용 프로그램의 가공된 데이터 시각화 (출력)담당
* 발표 및 문서화 보조 담당

**이상민**

* 파이썬 GUI, PyQt5 담당
* 주로 데이터 입력, 연산을 담당
* 온라인으로부터 응용프로그램으로 데이터입력 개발

**민성재**

- 파이썬 GUI, PyQt5 담당

- 버튼 설계 및 사용자 매뉴얼 구상, 에러처리

- 입력된 데이터를 맵에 표시 및 줌인/줌아웃

**정유석**

- 파이썬 GUI, PyQt5 담당

- Openstreetmap Api 적용 및 기본데이터 마커표시

- 버튼 설계 및 사용자 매뉴얼 구상, 에러처리 보조

# 주제의 필요성/중요성

많은 지역 기반 데이터가 있지만 입지, 특징 분석을 위해서는 사실상 사용자가 직접 분석을 해야 하는 경우가 많다. 지도의 위치만으로 보았을때 데이터의 분포가 어느 지역에서 어떻게 표현되는지는 단순히 데이터의 밀집도의 형태로 어느 동, 구에서 어느정도 집중되어있다라는 것만 알 수있다.

하지만 우리가 데이터를 분석할 때는 단순 집중도가 중요한 것이 아니라 주변 다른 요소들과 어떻게 상호작용을 하여 밀집이 되어있는지에 대한 인사이트를 얻고자 하기 때문이다. 예를 들자면 어떠한 건물의 가치는 건물들의 밀집도만으로 알 수 있는 것이 아니라, 주변의 도로와 같은 부수적인 요소가 중요하게 작용한다. 그렇기 때문에 **우리는 도로, 건물, 지하철 역의 분포를 베이스로 제공하여 데이터간의 상관관계를 분석하여 유저가 응용할 수 있도록 시각화, 정량화를 해주는 것으로 목표로 하고 있다.** 이때 건물이나 도로, 지하철 역의 정보는 서울시 공공데이터를 가공하여서 사용할 예정이다.

서울시 공공데이터는 대부분 .xls파일의 형태이고 그 데이터 또한 정규적이지 못하다. 단순히 정리만 해놓은 형태기 때문에 데이터셋간의 공통된 규칙이 존재하는것이 아니기에 실 사용과는 상대적으로 거리가 있다. 그렇기에 우리 프로젝트는 대표적으로 쓰일수있는 데이터셋을 정돈하고 개발자가 쉽게 사용할 수 있는 형태로 정규화 하는 것을 부수적 목표로 한다. 즉 우리는 사용자가 어떠한 데이터의 상대적 분포와 가치를 알고 싶을 때 유용하게 사용할 수 있는 툴을 개발할것이다.

이 프로그램은 오픈소스 커뮤니티에 기본이 되는 데이터 분석 툴을 제공하는 것이 목표이다. 기본이라 하면, 필요한 핵심기능이 있고 개발의 확장성이 높게끔 함수 설계와 재활용성을 고려하였고, 나아가 사용자의 입맛에 맞게 서비스로 활용할 수도 있으며 더 좋은 아이디어의 영감이 된다는 것을 정의로 한다.

# 관련 기술 소개

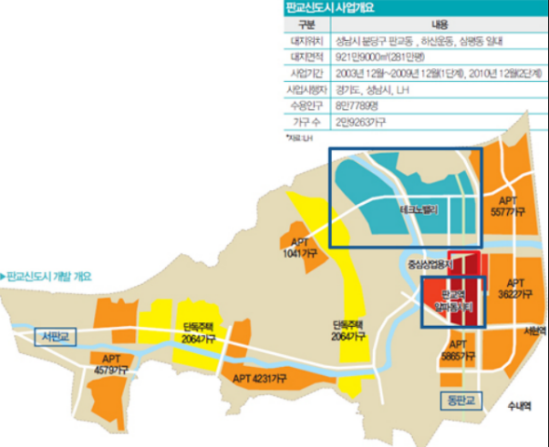
기존의 Opensource 프로그램들은 대부분 시각화 하는 부분에 초점이 맞춰져 있어 **단순 밀집도를 보여주는 알고리즘이 구현되어 있는 경우가 많다**. 또한 분석의 경우에는 직접 데이터의 분포와 형태를 보고 수기로 분석하는 경우가 대부분이고 지도 위에서 그런 데이터를 분석하는 과정은 찾기가 어려웠다

|  |  |  |
| --- | --- | --- |
| <https://github.com/e-/Multiclass-Density-Maps> | 다중 클래스의 밀집도를 보여주는 맵 | https://lh6.googleusercontent.com/X8LyzuG1jBk2h1oVCLp2MWvGRMJQT2nop79luYViQ5pJLtMrZPPkh_psdFmlbJeLtbCWBce8mItqX1kfJ0hmLIDo88JavGU_pOLMMt5px00oFs3Nu-qFxsdwojeRHYMGM18iTRA3 |
| <https://github.com/greatsong/2016dataproject> | 구 별 밀집도를 비교 |  |
| <https://github.com/CartoDB/cartodb> | 지역별 수에 따라서 색을 표현, 도로의 구조 등을 보여주며 유저가 modify할 수 있게 만든 플랫폼 | https://lh5.googleusercontent.com/FZ8hAlz57fRiJ-yGGImjRQSNuKbv4i4wnC9srT-J_6Yp9UKV8HqBs5hVPDInHeoV524PFxz3CGEhEfOuzrqxACOli3qS1E6xsmtSSXsDWuPFrAfXGOeqzvSMV_7yefuab5EH8Nn2 |
| <https://glasvase.tistory.com/14?category=672672> | 서울 공공데이터 통계 블로그 |  |
| <https://planet.openstreetmap.org/> | Openstreetapi에서 도로데이터 추출 |  |
| <https://github.com/ndrberna/CarAccidents> | 지도 데이터 시각화 |  |

1. 본인들만의 강점, 차별성  
    기존에도 지도를 활용한 여러가지 데이터 분석과, 그 데이터의 도식화를 위한 상용 소프트웨어는 있었다.  
    GIS[Geographic Information System ]는 일반 지도와 같은 지형 정보와 함께 지하시설물 등 관련 정보를 인공위성으로 수집, 컴퓨터로 작성해 검색, 분석할 수 있도록 한 복합적인 지리정보시스템이다. 기존의 GIS 프로그램은 크게 유료 S/W인 ArcGIS, Opensource인 QGIS가 있다.  
      
    오픈소스인 QGIS의 경우, 단순한 지도 유틸리티에 불과한 것이 아니라 여러 사람이 지도를 사용하여 작업한 여러 작업물들을 모아 놓은 하나의 아카이브이다. 그렇기에 지도데이터 이외에도 위성이나 등고선 등 여러 데이터셋 속에서 사용자가 원하는 데이터를 조합하여 지도위에 시각화하여 볼 수 있다. **하지만 여러 기능이 많은 QGIS의 경우에도 여러 기준 데이터들 위에서 시각화를 할 수 있지만 데이터 간의 분석 값을 얻기에는 힘들다.** **가장 큰 문제는 이러한 프로그램은 강좌를 듣거나 문서를 읽고 사용 방법을 배워야하는 진입장벽이 높은 프로그램이다**. 지도를 활용한 여러가지 절대적인 분석을 시각화 하는 기능 등 넓은 범용성을 갖추었지만 일반적인 사람들에게는 사용이 어려울 뿐더러, 무엇보다 데이터 분석을 위해서는 유저가 직접 설정을 하여 이용해야 하는 어려움이 있다.

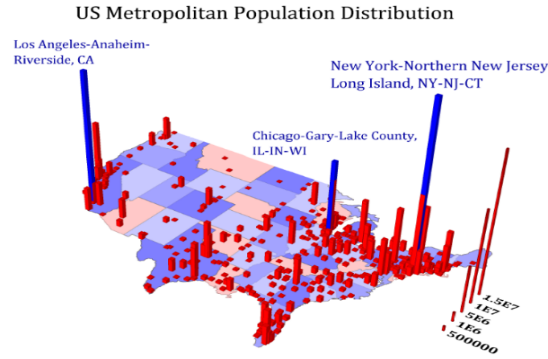
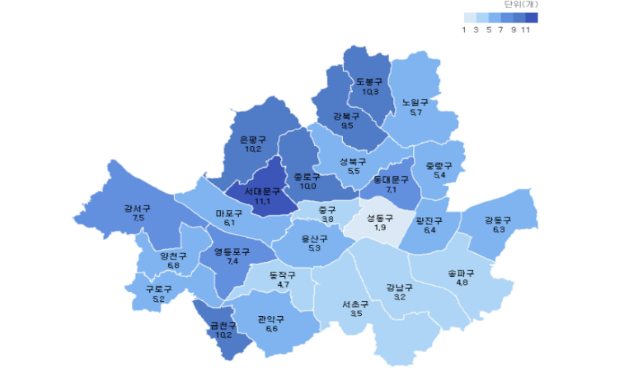
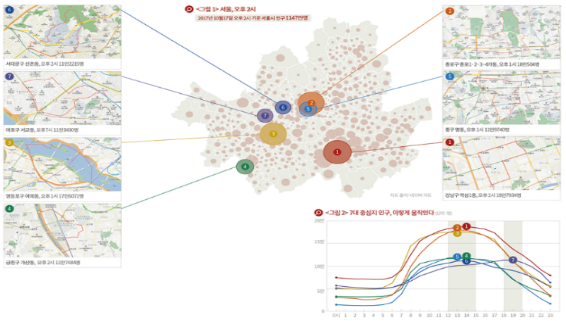


주요 도시들에 한해서는 자료들이 잘 조사되어 있지만 우리나라, 서울의 건물, 도로에 대한 분석은 부족한 점이 많고 오픈소스화 되어있지 않거나 빠져 있는 부분이 많다. 우리는 이러한 직접적인 데이터 분석의 과정을 없애고, 프로그램에서 사용자의 요구에 따라 서울 건물들의 가치를 직접 분석해서 보여주는 방식으로 일반 사용자들의 진입장벽을 낮추고자 한다.

거리를 걷다 보면 많은 광고들 에서 도로에서 특정 지점이 얼만큼 떨어져 있는지, 주변에 건물은 얼마나 많은 지, 고속도는 있는지, 주변에 지하철역이 있는지는 와 같은 지형적인 요소를 통해 본인들의 가치를 어필하곤 한다. 입지분석의 측면에서는 주변에 어떤 지형물이 있는지는 굉장히 중요한 요소이다. 입지분석을 넘어서 도시계획의 측면에서도 기존 데이터의 분석을 필수적이다. 단순히 가로등을 설치하는 문제에서도, 도로와 건물, 기존 가로등 사이의 비교를 통해 어떤 장소에 필요한지를 찾을 수 있다.

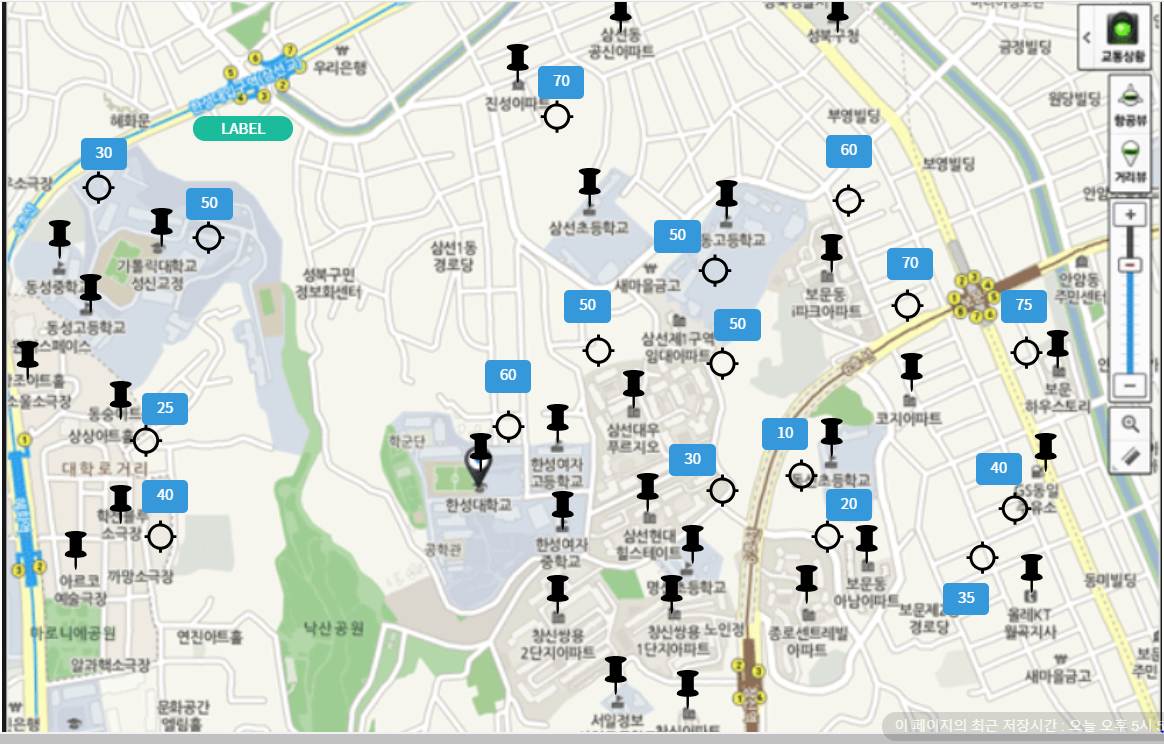
우리의 툴은 입지 분석에 초점을 두지 않고 사용자가 분석하고 싶은 데이터의 상대적인 분포와 가치를 분석하고 시각화 하여 제공하는 지도라는 점에서 독자적인 정체성을 갖는다. 도로, 건물, 지하철 역 등 우리 주변에 있는 여러 지형물들을 기준으로 데이터를 분석하게 된다면 사용자는 지역에 대해 보다 정확한 인사이트를 가질 수 있게 된다.

가치의 정도를 제시하는 방법 또한 주안점을 두고 있는 부분 중 하나이다. 데이터의 종류와 목적마다 계산하거나 초점을 두어야하는 정도가 각각 다를 것인데 이를 해결하기 위해 기존 논문에서 사용한 방법을 토대로 정량적으로 계산을 할 수 있는 방법을 찾을 예정이다.   
 **<GIS 기법을 활용한 은행입지분석에 관한 연구-이희연,김은미>**에서는 인구학적, 사회.경제적, 수요유발시설물들과 관련된 14개의 변수들간 중회귀분석을 하여 회귀모델을 추출하였다. 그 후 인구 분포와 비교하여 전체 지점에 대한 Candidate를 추출하여 최적입지를 찾아낸다. 하지만 우리는 각 지점의 가치를 계산해야 하기 때문에 회귀분석을 통해 각 인자 별 영향력을 추출해낸 후 분산과 표준편차를 통해 가치를 계산하고자 한다. 이는 관련연구를 참고하여 더욱 정확하게 진행할 것이다.   
  
  
  
  
 또한 우리는 사용자가 지도를 한눈에 이해하는 것에 초점을 맞추고 있다. 이에 대해 제시된 방법은 3차원 그래프, 다양한 색을 통한 마크업, 체크박스, 팝업 도표 등 여러가지가 논의되었으며, 타겟팅 한 사용자의 눈에 맞춰 데이터 시각화를 하고자 한다. 이때는 오픈소스로 공개되어 있는 여러 시각화 툴을 함께 사용해보고자 한다.



이 프로젝트는 상용화와 발전 가능성 뿐만 아니라Opensource의 측면에서도 여러 모듈을 추가하고 연결 할 수 있는 확장성이 있다는 점에서 주목할 만하다. 각종 시설, 지형물에 대한 지리적 가치를 제공 서비스는 편의적인 측면은 물론 상업적으로도 활용될 수 있기 때문에 사용자의 needs를 만족시킨다. 또한 프로젝트 진행 기간이 한정되어 지도를 서울로 국한하였지만, OpenSource이기 때문에 지도를 국내 전체로 확장할 수 있는 잠재력을 갖고 있다.

# 개발 방법 및 계획 (구현 범위, 사용할 tool등)





.

**우리의 Minimum Viable Product 는 다음과 같다.**

1. 우선, 위와 같은 Map GUI 응용프로그램 개발이 목표이다.
2. 사용자의 위치기반 (위도,경도)를 기반으로 한 데이터 값을 받아서 맵 위에 마커로 표시한다.
3. 사용자가 데이터의 종류나 중요하게 여기는 척도에 따라, 기존의 도로, 건물, 지하철 역의 weight값을 조절한다.
4. 사용자가 연산에 추가하고자 하는 기본데이터를 입력한다.
5. 사용자는 각 데이터 객체별로 주변의 기본 데이터들 과의 연관성을 찾기 위해 coverage를 입력할 수 있다.
6. 전체 데이터 객체마다 사용자가 지정한 coverage와 weight에 따라서 기본 데이터가 얼마나 포함되어 있는지 분포도를 계산하고, 이에 맞게 그들의 value를 측정한다.
7. 이를 ranking 혹은 3D 그래프로 시각화를 해준다

**추가적인 기능**

1. Zoom-In, Zoom-Out
2. 사용자가 입력한 데이터를 혹은 URL경로를 통해 해당 데이터를 프로그램이 입력한다.
3. 더 다양한 시각화 툴을 활용한다.
4. 단순히 입력데이터와 기본데이터의 분포도에 의한 가치계산이 아닌, 더 효율적이고 증명된 자료를 적용하기 위해 조사를 한다.

**개발 도구 및 방법**

개발 언어 : Python

개발 툴 : PyQt5 (Python GUI)

개발 형태 : 응용 프로그램

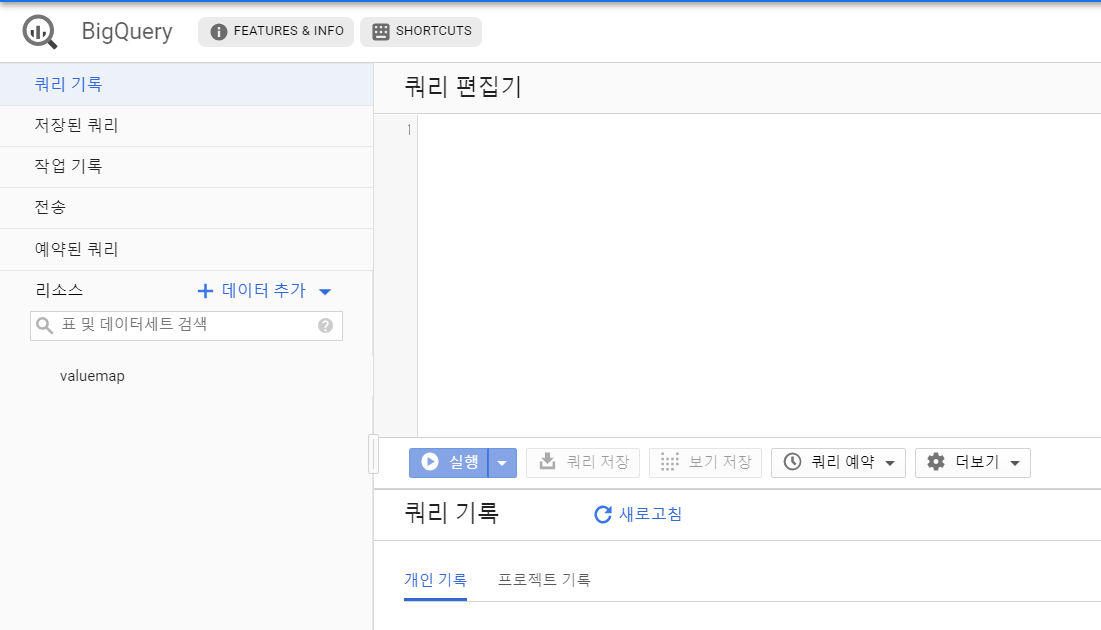
작업 방식 : Github

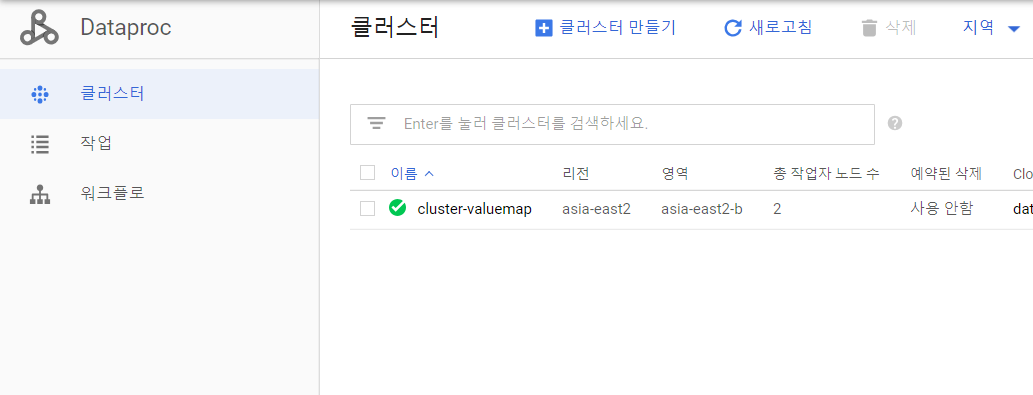
1. 개발기간이 시작되면, 전체적인 스케줄 및 매주 개인별 스케줄을 작성한다.
2. 개인별 스케줄에 맞춰 각자의 진행 상황을 Github에 남긴다.
3. 발표 및 매주 작업내역에 대한 문서화를 기록한다.

**- 개발 도구 옵션**

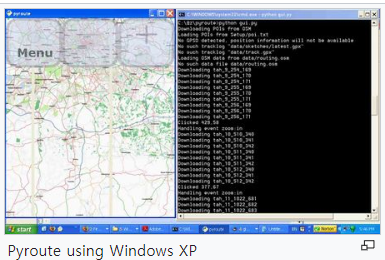
**옵션(1) <Google Cloud Platform-Dataproc/Bigquery**>

google에서 provisioning중인 GCP는 sql, 머신러닝, 서버, 데이터 분석등 많은 컴퓨터 자원이 필요한 여러 작업들을 하기에 최적화 되어있다. 위의 방식과 같이 전체적인 플로우를 팀에서 조직하는 것은 서버 구축, 배포에 드는 시간이 많을 것으로 생각해 이을 줄이기 위해 우리 팀은 구글 클라우드 플랫폼 위에서 서비스를 진행하는 방법도 고려 중이다.





유저가 보낸 데이터셋을 gcp의 Dataproc이나 Bigquery에서 받아 우리가 가지고 있는 도로, 건물 과 같은 공공데이터셋과의 상관분석을 통해 각 데이터의 상대적 가치, 지리적 중요도를 알아내고자 한다. 다중 서버로 동시에 처리를 한 후 이를 주 서버로 보내 응답 시간, 처리 방식을 간단하게 하고자 한다.   
  
  
**옵션(2) <Opensourcemap GUI/ Python GUI** >

<https://wiki.openstreetmap.org/wiki/Pyroute>  
우리의 작업이 Opensource적으로 사용되거나 지속적인 업데이트, 개인 유저가 사용하기 위해 서버리스의 형태로 가는 방식도 고려 중이다. 오른쪽에서 보이는 것 과 같은 형태로 로컬에서 실행하게 되면 우리가 제작한 분석 알고리즘이 유저의 컴퓨터에서 연산을 하여 보다 쉽게 사용할 수 있는 방식이다. 이를 위해서는 데이터 연산 과정을 최적화하는 과정이 필요할 것으로 생각된다.   
 로컬에서 실행할 수 있게 하기 위해서 다양한 포트와 확장성을 목표로 작업할 것이다. 또한 기존 GIS 플랫폼과 연계가 될 수 있도록 하는 것을 부수적인 목표로 할 예정이다. 전체적인 구성은 상단에서 작성한 방식으로 진행되며 단지 그 표현 방식이 우리가 maintain하지 않아도 지속적으로 사용가능 하도록 구성하는 것으로 바뀐 것이다.