

회의 데이터

Q: 사분위수로 서울지역인지 확인해도 되는가?

A: 사분위 수로 해도 상관없다.(데이터 분석하는 사람들의 마음이다.)

---

Q: R로 SHP파일 변환해도 상관없는가?

A: 마음대로 해라

---

Q: Line string값만 있는 파일을 어떻게 처리해야 할지 모르겠다. 그리고 왜 이런 일이 발생했는지 모르겠다.

A: 녹색교통 개발자와 미팅 협의해서 해결하자.

A: 1년치 데이터 볼 때, Line string 값만 있는 거의 비율을 추측해봐라

---

Q: 동 단위보다 더 작은 단위의 데이터를 얻을 수 있는가?

A: 줄 수 있다. 하지만 그게 왜 필요한지와 어떤 단위를 쓰는 게 좋은지 EDA가 필요하다.

줄 수 있는 자료:

서울시가 갖고 있는 자료: 25(자치구) / 424(행정동)

통계청에서 가지고 있는 자료: 6,000(집계구) / 80,000(조사구) / 160,000(50 by 50)

주무관님은 작은 단위를 쓸거면 기초단위구역(4,000)을 추천하셨습니다.

---

Q: 유사도 계산을 어떤 방식으로 해야 할지 모르겠습니다.

A: 유사도를 계산할 때 device 아이디가 있으면 굳이 할 필요가 없다. Device 아이디를 확인할 수 있는지 알아보고 이걸 알 수 있으면 굳이 할 필요 없다.

---

Q: 이동평균을 잘 모르겠다.

A: 이동평균이 좋을지, 커널 방식이 좋을지는 좀 더 생각해봐야 한다. 근데 이동평균에 대한 개념을 알고 있으면 뭐 딱히 알려줄 게 없다.

A: Time Interval에 대한 것만 알게 되면 쳐내는 것에 대해 판단할 수 있지 않나?

Time Interval에 대한 부분은 좀 더 알아보겠다.

---

Q: Moving distance 값을 어떻게 해결해야 할지 모르겠다.

A: 두 가지 방법을 다 해봐라. Start – end 포인트랑 Quantile의 직선거리 구해서 값을 채우거나 또는 파일의 용량에 따라서 장거리, 단거리를 구분할 수 있는지도 봐달라.

덧: 앱을 수동적으로 키는 것이기 때문에 유저의 행동패턴에 대한 로그와 같다. 이 사람들이 앱을 언제 어떻게 쓰는지를 모를 경우가 많다. 이 사람들이 언제 어떻게 키는지 패턴도 볼 수 있으면 좋겠다.

덧2 : 작업과정을 수시로 기록해서 주무관에게 공유해달라

덧3: 깃허브 주소 알려달라

덧4: QPY 사용할꺼면 stack overflow에 구문 있으니 찾아서 써도 된다.

---