

기업연계 멘토링 진행

| | | | |
|-----|---------------|-------|------------|
| 차수명 | 2차시 | 일자/시간 | 2021-10-21 |
| 팀장명 | 김예찬 | 기업명 | 신한은행 |
| 주제 | 설계 점검 및 기타 질문 | | |

세부 내용

[참여인원]

5명 전원 참석 (김예찬, 김윤서, 방지환, 이규정, 정희진)

[주요 진행 내용]

- 아키텍처/프로토타입 피드백 수렴
- 데이터 분석 기준 검토
- 그 외 데이터 처리 방법 논의

[질문/답변 내용]

- 분석 데이터를 CSV 형식으로 뽑아낸 후 분석하는 것인가요?

보통 행내 데이터는 DW에 웬만하면 들어있지만, 현재는 용량과 방법에 한계를 느끼는 중입니다. 최근 DB에 올리지 않고 파일 그대로 받아서 사용하는 경우가 다수입니다.

- 이상치 제거 전/후 그래프를 비교 분석하는 것이 좋을까요?

수치형은 이상치 제거, 범주형은 null 값 제거를 통해 나온 결과를 반응형으로 보여주었으면 좋겠습니다. 특히, null 값을 모두 지우면 결과에 큰 영향이 갈 수 있기 때문에 그것을 주의하세요.

- 아키텍처 및 figma 피드백

Figma의 boxplot을 잘 설정해주셨습니다. IQR말고도 영향을 줄 수 있는 변수를 많이 찾아서 적용해보세요. 아키텍처와 기능명세에는 특이사항은 없습니다.

- Feature 후보군 중 어떤 것을 선호하시나요?

저정도 feature면 충분할 것 같습니다.

- 행내 이상치 처리 분류 기준은?

현재 존재하지 않아서 자유롭게 만들어주시면 됩니다.

[멘토 전달사항]

★ 핵심 - 수치형의 outlier, 혹은 범주형의 null 값 제거 시 다른 값들이 영향 받아서 통계치가 바로 바뀌는 모습을 보여주기

- string - id처럼 개별 값인지, 범주형에 속할 수 있는지 **auto detect**가 되어야 할 것

- 수치형 + 범주형 경우, 기계적으로 min, max 계산이 아닌 다른 방법 써야 합니다.

- test (dummy) 데이터 확인 완료

[건의사항]