**[붙임 2] 2025 데이터분석 연구발표회 계획서 제출 양식**

| **2025 데이터분석 연구발표회 연구계획서** | | | | |
| --- | --- | --- | --- | --- |
| 팀명 | 지팡이 | | | |
| 참가자 명단  (3명 이내) | 학번 | 20802 | 20924 | 21007 |
| 이름 | 권민승 | 최서윤 | 김인하 |
| 1. 문제 정의  (연구 주제) | 고령화에 따른 디지털 격차와 사이버 범죄 발생율의 상관관계 | | | |
| 2. 데이터 수집 | 필요한 데이터의 요건 및 이용할 사이트, 자료 수집 방법을 기재함  **사이버범죄율**:   * 경찰청 공공데이터 개방 → 2021 시도청별 사이버범죄 발생 및 검거 데이터 * 전자기기 보급율: * 온라인 상거래 이용률: * 디지털 교육: 공공데이터 - 지역별 디지털 배움터 이용자 수   **고령화**:   * 주민등록 인구통계 → 행정구역별 평균 연령 데이터 → 고령화 정도 수치화 * 독거노인 가구비율: KOSIS 통계청 e-지방지표 [독거노인가구비율](https://kosis.kr/visual/eRegionJipyo/themaJipyo/eRegionJipyoThemaJipyoView.do?themaId=A_01_02&menuThemaId=A_01_02_02&jipyoId=5440_6763&jipyoNm=&graphTypeGbn=THEMA&statId=&regionChkVal=00%40&chartGbn=DTypeChart&selectPrdDe=&themaGbn=subjectJipyo&detailJipyoId=&themaGbnMenu=subjectJipyo&chooseYm=&jipyo1PrdDe=04bcd4bcd&AreaChoiceCombo=A_01_02) , 공공데이터포털 * 평균 학력: 교육통계서비스 KESS | | | |
| 3. 데이터 전처리 | * 중복값을 제거하거나 데이터 통합 및 구조 변경에 사용할 툴을 선정하고 방법을 소개함 * 중복값 제거, 이상치 및 결측치 보정, 데이터 연계/통합, 데이터 구조 변경 * 비교할 지역 선정, (고령화율이 높고 사이버범죄 발생율이 높은 지역, 높은 고령화율+낮은 사이버범죄 발생율, 낮은 고령화율+높은 사이버범죄 발생율, 낮은 고령화율+낮은 사이버범죄 발생율)-> 클러스터링(k-means 방법 이용) * 오렌지3: python script를 활용해서 고령화율과 사이버범죄 발생율, 전자기기 보급률, 학력 수준 등의 데이터 연계/통합 | | | |
| 4. 데이터 모델링 | * 데이터 간의 관계 설정에 사용할 툴과 방법을 소개함 | | | |
| 5. 시각화 및 탐색 | * 시각화에 사용할 툴과 방법을 간단히 소개함 * 시각화는 flourish를 활용해서 우리나라 지도의 사이버 범죄율, 고령화 인구 비율 등을 나타냄   사이버범죄 발생율과 독거노인 가구비율을 Geo라는 addon을 활용하여 지도형 이미지로 시각화하여 비교하고,  관련 데이터인 전자기기 보급율과 디지털 교육, 평균 학력 데이터를 오렌지를 통해 변환하여 관계성 파악    지역별 고령자 수 / 총 인구수를 통해 지역별 고령자 비율을 수치화하여 flourish 시각화: | | | |