**[ 6월 3주차 프로젝트 수행 일지 ]**

**프로젝트 수행 기간 : 6/9 ~ 6/16**

| **프로젝트 주제** | **서울특별시 내 시니어들의 특성에 따른 문화생활 참여 예측 및 소비 분석** | | |
| --- | --- | --- | --- |
| **프로젝트 팀명** | **김家네** | **프로젝트 팀원** | **김종원, 김지우, 김연진, 김예리** |

**※**

| **1. 프로젝트 수행 계획 및 현황** | | |
| --- | --- | --- |
| **이름** | **금주 프로젝트 수행 내용** | **차주 프로젝트 수행 계획** |
| **김종원** | 1. **주제 아이디어 내기**  * **지역별 청소년 체육 지원 시설 및 프로그램 개수에 따른 청소년 체육 활동 빈도 분석**  1. **주제 관련 리서치**  * **시니어들의 관심사 및 여가 생활 현황 조사** * **현재 여가 생활 여건에 대한 엑티브 시니어의 만족도 파악**  1. **데이터 수집**  * **모델에 대해 추가 feature를 위한 데이터 수집** * **메인 데이터와 서브 데이터 병합 및 크기, feature 개수 점검**  1. **주제발표 자료 제작** 2. **주제 및 데이터 점검 회의** | **19-20일 전처리**   * **전처리 진행 방향 논의** * **데이터 정제 및 병합** * **이상치 제거**   **21-23 시각화, 분석, 머신러닝**   * **분석 시각화 분담 후 수행** * **머신러닝 알고리즘 별 모델링** |
| **김지우** | 1. **주제 아이디어 내기**  * **아이 동행 안심 나들이** * **창작도우미** * **노인요양병원 리뷰 커뮤니티 플랫폼** * **지역별 / 나이대별 문화 소비 현황 분석**  1. **주제 관련 리서치**  * **문화분야별 참여도 분석** * **시니어 소비 트렌드 분석** * **아이랑 함께 할 수 있는 문화생활 리서치**  1. **데이터 수집문화**  * **구입 문화콘텐츠 종류** * **서울특별시 문화소비 평균비용 통계** * **여가문화 유형별 관심도** * **아이랑 문화생활 정보** * **행정구역별 인구대비 문화생활 정보**  1. **주제발표 자료 제작** 2. **주제 및 데이터 점검 회의** | **19-20일 전처리**   * **전처리 진행 방향 논의** * **데이터 정제 및 병합** * **이상치 제거**   **21-23 시각화, 분석, 머신러닝**   * **분석 시각화 분담 후 수행** * **머신러닝 알고리즘 별 모델링** |
| **김연진** | 1. **주제 아이디어 내기**  * **서울시민의 독서량에 영향을 미치는 요인 분석: 자치구별 도서관 보급률, 자치구별 인구 특성, 도서관별 도서 예산, 도서관 소장 목록, 유동 인구 특성 등으로 나누어 살펴보고, 자치구별 독서 활동 증진을 위한 전략 모색**  1. **주제 관련 리서치**  * **시니어 세대의 인구 동향 조사** * **시니어 세대의 경제적 특징 조사**  1. **데이터 수집**  * **2021 년 장래인구추계를 반영한 내외국인 인구전망 [통계청]** * **삼정 인사이트[삼정 KPMG 경제연구원]** * **2022 고령자 통계** * **성 및 연령별 추계인구 (1 세별, 5 세별) /전국** * **거주지역/가구주의 성/연령대별 주택소유 가구수** * **가구주 연령별 가구당 월평균 가계수지** * **문화체육관광부\_전국공공도서관통계** * **국내 문화체육관광 분야 공공도서관 시설 및 소장자료 통계 데이터(이용건수 등 포함)**  1. **주제 발표 자료 제작** 2. **주제 및 데이터 점검 회의** | **19-20일 전처리**   * **전처리 진행 방향 논의** * **데이터 정제 및 병합** * **이상치 제거**   **21-23 시각화, 분석, 머신러닝**   * **분석 시각화 분담 후 수행** * **머신러닝 알고리즘 별 모델링 분담하여 수행** |
| **김예리** | 1. **주제 아이디어 내기**  * **연령층별 공연 선호도 분석을 통한 추천 서비스** * **국민체력증진을 위해 건강정보 분석 후 체력 시설, 스포츠강좌 알림 서비스**  1. **주제 관련 리서치**  * **연령대별 시니어들의 문화참여도가 늘어나고 있는가?**   + **연령대별 문화예술관람률**   + **연령대별 여가시간**   + **국민여가활동 조사**   + **시니어의 문화 여가활동 유형조사**  1. **데이터 수집**  * **연령층별 공연 선호도 분석을 통한 추천 서비스**   + **공연예술통합전산망**   + **예술의 전당**   + **정동극장**   + **부산문화회관 문화전시 성별 연령별 이용현황 데이터** * **국민체력증진을 위해 건강정보 분석 후 체력 시설, 스포츠강좌 알림 서비스**   + **국민체력측정 현황 데이터**   + **체력측정별 운동처방 데이터**   + **체육시설알리미**   + **스포츠강좌이용권**   + **국민체력 100사업**  1. **주제 발표 자료 제작** 2. **주제 및 데이터 점검 회의** | **19-20일 전처리**   * **전처리 진행 방향 논의** * **데이터 정제 및 병합** * **이상치 제거**   **21-23 시각화, 분석, 머신러닝**   * **분석 시각화 분담 후 수행** * **머신러닝 알고리즘 별 모델링 분담하여 수행** |

| **2. 강사 피드백** | |
| --- | --- |
| **유선준** | 1. **주제 관련 피드백**  * **아이 동행 안심 나들이 (폐기)**   + **출산율이 급격히 저하되는 현 상황에 아이와 관련된 주제로 나아가는게 사업성이 있을까? 공모전 주최 측에 입장에서는 크게 와닿지 않을 수도 있다고 말씀해주심.**   + **지역 축제, 관광 명소 이렇게 애매하고 너무 넓은 주제 말고 아이들이 갈만한 곳, 예를 들어 유원지, 놀이공원, 동물원, 체험학습장 등 이런 곳을 타겟으로 정하는 게 좋겠다고 추천해주심. 그리고 반대로 아이들이랑은 잘 안가려고 하는 곳도 분석해서 비교를 해보는 것도 좋겠다고 의견을 주심.** * **지역별 / 나이대별 문화 소비 현황 분석**   + **분석하기에는 좋지만, 인사이트 도출 및 사업성 부문에서 어떻게 해야할지 모르겠다고 강사님께 여쭤봄.**   + **여러가지 방향성을 제시해주셨고, 일단 양질의 빅데이터를 찾을 수 있는지, 데이터에 초점을 맞춰서 생각해야한다고 피드백을 주심.**   + **이후에 지역별 / 나이대별 에서 ‘시니어'로 주제를 좁혀서 가는 방안에 대해 말씀드렸고 좋다고 해주심.**  1. **분석 방향 관련 피드백**  * **기존에 머신러닝 기법으로 군집화를 사용하려고 했었는데, 군집화는 무언가를 예측할 때 쓰는 알고리즘이 아니라 기존에 있는 데이터 파악 정도의 용도로 쓰는 알고리즘이라고 말씀해주심.** * **주제를 정하기 전에 어떤 기준을 가지고, 왜 이 주제를 정하게 됐는지 이유가 필요하다. 그를 위해서는 사전 분석을 해야할 필요성이 있다고 말씀해주심. 기존 트렌드 분석이라던지, 관련된 조사, 보고서, 또는 논문을 본다던지 등등**  1. **수집한 데이터 관련 피드백**  * **기존에 구했던 ‘구입 문화디지털 콘텐츠 종류‘ 데이터는 설문조사 데이터인데, 이 데이터를 활용하면 어떤 단점들이 있는지 설명해주심. 그리고 이 데이터를 이용하여 특정 머신러닝 기법을 응용하는 것이 얼마나 유의미한지에 대해 피드백을 주심.** * **이 데이터를 가지고 뭔가 분석을 하는데 feature 값이 너무 적어서 의미가 없다는 피드백을 주심.** * **‘공연전시체험구매여부’를 target으로 설정할 경우, 예측된 target이 의미가 없으며, 수치형 데이터가 필요하다는 피드백 받음**   + **새로운 데이터 셋을 통해 수치형 데이터인 ‘문화예술행사관람실태직접관람횟수’를 target으로 설정**   + **단순히 참여 여부만을 알 수 있기 때문에 빈도수에 관해서는 알 수 없다는 피드백을 주심. yes 값이 많다고 해서 참여율이 많다고 할 수 있는지 근거를 제시할 수 있는 데이터가 필요하다고 말씀해주심.**   + **위에 이어서 문화 생활 관련된 수치 데이터가 있는지, 없다면 수집이 필요하다고 말씀해주심.** |