**[ 6월 5주차 프로젝트 수행 일지 ]**

**프로젝트 수행 기간 : 6/24 ~ 6/27**

| **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. **머신러닝 모델링**  * **개인의 특성에 따른 ‘문화예술행사 참여 횟수’의 변화 분석** * **로지스틱 회귀분석 응용** * **하이퍼 파라미터 설정**  1. **머신러닝 결과를 토대로 소비분석**  * **가구 소득에 따른 문화 활동 소비 분석** * **시각화 및 분석**  1. **발표 자료 작성** 2. **발표** |  |
| **김지우** | 1. **국민문화예술활동조사 시각화 및 분석**  * **기존에 K-Modes 알고리즘을 통한 군집화 말고 K-Means와 K-Prototypes 군집화도 같이 진행** * **시니어에 해당하는 워드클라우드 시각화 추가** * **군집화 및 워드클라우드 발표 자료 추가**  1. **머신러닝 모델링**  * **랜덤포레스트 분류 진행** * **머신러닝을 위한 전처리 완료** * **하이퍼 파라미터 튜닝** * **모델 성능 향상을 위해 범주 축소 및 데이터 늘리기 진행** * **피쳐 중요도 시각화**  1. **소비분석 진행**  * **소비분석 어떤 방향으로 진행해야 할 지, 어떤 걸 도출해내고 싶은지 회의** * **기존의 머신러닝 했던 결과를 어떻게 소비분석과 연계할지 고민** * **소비 분석 인사이트 도출**  1. **발표자료 만들기**  * **각자 분담해서 진행했던 부분 발표 자료 추가** * **모든 발표자료 취합한 이후 실제 발표에 쓸 ppt 추리기** |  |
| **김연진** | 1. **머신러닝 모델링**  * **데이터 전처리** * **타겟 데이터 범주형 변환** * **LightGBM classifier 학습 및 예측** * **하이퍼파라미터 튜닝** * **피쳐 중요도 시각화**  1. **소비 분석 데이터 시각화**  * **소비 분석 데이터 EDA** * **상위/평균/하위 관람 그룹 관람 걸림돌 분석** * **타겟 그룹 분류 후 기본 분석**  1. **발표자료 제작**  * **발표 내용 정리** * **발표자료 시각화** |  |
| **김예리** | 1. **서울시 문화활동 연간 평균비용 통계 데이터 전처리**  * **column 정의** * **범주형 코드값 통일** * **파생변수 생성** * **데이터 병합**  1. **데이터 시각화**  * **국민문화예술활동조사 데이터에 대해 필요한 부분 시각화**  1. **발표자료 제작 및 마무리** |  |

| **2. 강사 피드백** | |
| --- | --- |
| **유선준** | 1. **분석 방향 관련 피드백**  * **군집화**   + **군집화를 한 이유에 대해서 설명이 필요하다고 말해주심**   + **시니어 데이터를 토대로 군집화 하여 군집 별로 특성 파악해야한다고 말씀해주심**     - **특성 파악으로는 income 열을 해보는 것이 좋겠다고 말씀해주셨으나, 유의미한 결과가 도출되지 않아서 expense열로 특성 파악 후 시각화 진행하였음.**   + **최종적으로 군집화를 진행하는데 쓴 알고리즘은 K-Prototypes 하나만 쓰는 것이 좋을 지 아니면 다른 알고리즘 K-Modes 도 발표 자료에 넣어야 할 지 여쭤봤는데 세 가지 모두 진행해본 결과 K-Prototypes 알고리즘이 가장 좋게 나와서 그것에 대한 결과를 말한다고 이야기하면 좋을 것 같다고 피드백 해주심.**  1. **발표 방향성 관련 피드백**  * **도출한 결과에 대해 과정 전체가 아닌 간단하게 보여주고 어떻게 도출해냈는지에 대한 과정을 포함을 말씀해주심**   + **중간 결과는 한 번씩 짚고 넘어가기**   + **하나의 분석과 알고리즘을 적용했을 때 무엇을 얻었는지 짚고 넘어가기**   + **분석 내용 바탕으로 예측결과 제시하기** * **머신러닝 결과 발표 방향성**   + **무엇을 얻고 싶어서 머신러닝으로 예측한 것인지 내용 포함하기**   + **무엇을 알고자 분류를 하고 회귀를 했는지 각 feature와 연관시켜 설명하기**   **왜 분류를 더 많이했는지에 대한 내용 포함하기**   * **소비분석 발표 방향성**   + **무엇을 알고자 소비분석 데이터를 사용했는지 포함하기** |