|  |  |
| --- | --- |
| 교육 제목 | **시각화 도구와 데이터 사전 처리** |
| 교육 일시 | 2021.11.08 |
| 교육 장소 | 영우 글로벌 러닝 |
| **교육 내용** | |
| 오전 | 1. **시각화 도구**   -seaborn 라이브러리  Histplot-히스토그램,kdeplot-커널밀도함수,heatmap-히트맵  Strip plot-데이터 분산 미고려, swarmplot-데이터 분산 고려  Barplot-막대 그래프,countplot-빈도 그래프,boxplot-박스 플롯  Viloinplot-바이올린 그래프,jointplot-조인트 그래프  -folium 라이브러리  Folium.Map(location=[위도,경도],zoom\_start=줌위치)   1. 데이터 사전처리   -누락 데이터 처리  Isnull과notnull로 누락데이터를 찾는다  Dropna로 누락 데이터를 제거  Fillna로 누락 데이터를 치환  -중복 데이터 처리  Duplicated method를 이용하여 중복을 찾는다.  Drop-duplicates()-중복을 제거  -데이터 표준화  단위 환산시 새로운 칼럼을 만들어 환산한 Series를 붙인다.  Astype을 이용하여 원하는 데이터 자료형으로 변환한다. |
| 오후 | 1. **데이터 사전처리**   -범주형 데이터 처리  히스토그램을 이용하여서 구간을 분할한다.  Pd.cut을 이용하여서 각 구간에 데이터들을 분할한다.  Pd.cut(데이터 소스,경계값 리스트,bin 이름, 첫 경계값 포함)  -더미 변수  컴퓨터가 인식 가능하기 위해 사용하는 변수로 해당 특성의 여부를 표시한다.  -정규화  숫자 데이터의 상대적 크기 차이를 줄이기 위한 수단으로 정규화를 사용하는데 이경우 데이터의 범위가 0~1혹은 -1~1이 된다.  -시계열 데이터  각 데이터를 시간순에 따라서 나열하여 한 눈에 볼 필요가 있을 때 사용하는 정리기법으로 사용한다. |