|  |
| --- |
| **SSAC**  **파이썬을 활용한 비즈니스 빅데이터 분석가 양성 과정**  **(TEMPLATE) 자가학습 워크시트** |

|  |  |
| --- | --- |
| **이름** | 백승재 |
| **날짜** | 2021-10-18 |
| **온라인 콘텐츠(인강)**  **학습 범위** | * 직장인을 위한 Python : CH\_01\_11 ~ 전처리\_03 |
| **복습 노트 작성 (자유 양식)** | Ch\_01-11. Row,column 추가 및 삭제  df.append({key:value,ignore\_index=True}, df.loc[df[column name]==’’,’’]=’value’  Ch\_01-12. 통계값(min,max,mean, var,std,count,sum,median,mode)  Std= np.sqrt((Df[i]-Df.mean())\*\*2 ….+ Df[n]-Df.mean())\*\*2)/n)  Ch\_01-13. 피벗테이블(pivot\_table)  pd.pivot\_table(df, index=’행 인덱스’, columns=’열 인덱스’, values=’조회하고 싶은 값’, aggfunc = np. 함수)  Ch\_01-14. 그룹별 통계(group by) 함께 사용되는 함수 (sum,mean, count,min,max,var,std)  Ch\_01-15. 복합 인덱스(multi-index)와 인덱스 초기화 (reset\_index)  복합 인덱스df.groupby([]) -> Unstack() -> 인덱스 초기화 (reset\_index)  Ch-전처리-01.결측값 채우기 (fillna)  데이터에 결측값 유지 : inplace, fillna값을 넣어주는 방식, 결측값 처리: 0,평균값(mean), 최빈값(mode), 중앙값(median)  Ch-전처리-02.결측값 제거 (dropna), 중복제거(drop\_duplicates)  df.dropna(axis=1, how=”any(or all)”), df.drop\_duplicates(“column 명”,first(or last))  Ch-전처리-03.행과 열 제거 (drop)  df.drop(‘column명’, axis =1), df.drop([‘column명’,’column명’], axis =1), df.drop(‘row명’,axis=0) |
| **퀴즈 제출** | Q1. Data 01의 데이터 분포를 표준편차를 통해 확인하려고 한다. 표준 편차를 구하는 과정에 해당하지 않은 것은?  data\_01 = np.array([1, 3, 5, 7,8,10,10,11,12,14])  1) np.sqrt(((1-8.1)\*\*2 + (3-8.1)\*\*2 + (5-8.1)\*\*2 + (7-8.1)\*\*2 + (8-8.1)\*\*2+(10-8.1)\*\*2\*2+(11-8.1)\*\*2+(12-8.1)\*\*2+(14-8.1)\*\*2 ) / 10)  2) data\_01.std()  3) data\_01.mode()  4) data\_01.var()  정답 : 3  Mode는 최빈값을 구하는 함수로 dataframe에서 사용가능 하다.    Q2. DF데이터에서 혈액형별 키에 대한 평균값을 출력한 것이 아닌 것은?  1) pd.pivot\_table(df,columns = "혈액형", values="키",aggfunc=np.mean)  2) df.groupby(“혈액형”)[‘키’].mean()  3) df.loc[:,['혈액형','키']].mean()  4) df.groupby(["혈액형"]).mean().reset\_index()  정답 : 3  - 3번은 혈액형과 관계없이 모든 키의 평균을 출력한다.    Q3.다음 데이터 프레임에 ’키’열에 결측값을 처리해 데이터프레임(원본 또는 복사본)에 적용하는 방법으로 적절하지 않는 것은?  #df1~4는 df의 복사본이다.  1) df1['키']=df1['키'].fillna(0)  2) df2['키'].fillna(df['키'].mean(),inplace=True)  3) df['키'].fillna(df['키'].median())  4) df3['키']=df3['키'].fillna(df['키'].max())  정답 : 3 원본에 적용이 안되며 변경 시 inplace옵션 또는 fillna을 이용해 값을 넣어줘야 한다.  Q4. 데이터 프레임에 키에 대한 중복값을 가지는 행을 제거(first 중복값 제거)하는 전처리 구문으로 적절한 것은 ?  1) df['키'] = df['키'].drop\_duplicates(keep='first')  2) df.drop\_duplicates('키', keep='last')  3) df['키'] = df['키'].drop\_duplicates(keep='last')  4) df.drop\_duplicates('키', keep='first')  정답 : 2 (뷔(3)과 소연(7) 행 제거됨)  - 1. 정국 키 NAN 처리됨  - 3. 뷔 키값 NAN으로 변경됨  - 4. 정국(5)과 태연(10) 행 제거됨 |
| **교육생 사진 (교육생 모습 + 교육 화면)** |  |
| **화면 캡쳐**  **(교육 화면 + 교육일시)** |  |