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

|  |  |
| --- | --- |
| **이름** | 백승재 |
| **날짜** | 2021-10-11 |
| **온라인 콘텐츠(인강)**  **학습 범위** | * 직장인을 위한 Python : CH\_01\_01 ~ 01\_10 * Python : CH08-01 ~08-04 |
| **복습 노트 작성 (자유 양식)** | CH\_01 : Pandas 소개 ~ 01-02 Series와 dataframe  pd.**S**eries([]), pd.**D**ata**F**rame(), df1.columns[],df1.index=df1[]  CH\_01-03: CSV 파일 읽어오기 pd.read\_csv()  CH\_01-04: Excel 파일 읽어오기 pd.read\_excel()  Ch\_01-05: 통계값(describe), 요약정보(info), 정렬(sort)  df.index, df.columns, df.describe(), df.info(), df.tale(),df.tale() df.sort\_index(), df.sort\_values(by=[‘키’,’브랜드평판지수’,ascending=True]  Ch\_01-06. Selection-column 선택 : df.[‘컬럼명’]  Ch\_01-07. Selection-loc,iloc : loc[,], loc[,[]], iloc[,], iat[,]  Ch\_01-08. Boolean Indexing : 비교 연산자 사용, isin(), iloc 활용 -> Boolean 값 -> fancy indexing  Ch\_01-09. 결측값 색인 – isnull, isna. Notnull  Info() -> Isnull(), isna() -> 결측값 확인(isnull fancy indexing) -> notnull 확인 -> loc fancy indexing  Ch\_01-10. 복사 (copy) df\_copy = df.copy()  CH08-01. 데이터 소개와 분석 목표 설명  - kaggle 2020 Machine Learning & Data Scienece Survey  - Kaggler의 교육 상태 확인, Kaggler의 Job Position 확인  CH08-02. 문제풀이 1  - 범주별로 리스트를 만듦-> 각 범주마다 필요컬럼 선택 -> dataframe으로 병합  - 결측치 있는 행 확인 isnull함수, isnull함수.any(axis =1)  - 결측치 행 제거 : drop na함수  CH08-03. 문제풀이 2 그래프 시각화(barplot, histplot, piplot)  CH08-04. 문제풀이 3. 데이터 정렬 후 그래프 그리, 데이터 컬럼을 isin함수 사용하기 |
| **퀴즈 제출** | Q1. DF를 이용하여 특정값을 선택한 것중 출력값이 다른 하나는?    1) df.loc[:1,'혈액형']  2) df.loc[:1,['혈액형']]  3) df.iat[2,6]  4) df.iloc[:2,6]  정답 : 3 iat 함수는 특정 위치의 값 1개만 출력한다.  1,2,4 출력값 3 출력값    # 1,2의 출력값은 동일하지만 Dtype은 다르다. (.1 series, 2.dataframe)  Q2.DF를 이용하여data Selection을 하려고 한다. 적절하지 않는 것은?  1) df.loc[ df['키'] > 180, ['이름', '키']]  2) df[df['키']>180]['이름','키']  3) df[df['키']>180]['이름']  4) df.loc[ df['소속사'].isin(my\_condition), '소속사']  정답 : 2 실행시 keyError 발생    2: ) []로 컬럼 리스트로 감싸준다. -> df[df['키']>180][['이름','키']]  Q3. 다음 dataframe 중 값이 다른 것은?    df\_new1 = df  df\_new2 =df\_new1  df\_copy=df\_new1.copy()  df\_new1[‘이름’] = 0  1) df\_new1  2) df\_new2  3) df  4) df\_copy  정답 4:  1~3번은 연결되어 있으며 코드 4번 줄의 인덱싱으로 인해 ‘이름’ 컬럼이 0으로 바꿨다.  4번은 사본으로 1~3번과 독립된 DataFrame이다. |
| **교육생 사진 (교육생 모습 + 교육 화면)** |  |
| **화면 캡쳐**  **(교육 화면 + 교육일시)** |  |