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

|  |  |
| --- | --- |
| **이름** | 백승재 |
| **날짜** | 2021-09-27 |
| **온라인 콘텐츠(인강)**  **학습 범위** | * 직장인을 위한 Python : 0502~0506 * Python : Numpy 06\_02 ~ 06\_07 |
| **복습 노트 작성 (자유 양식)** | 05-02.함수 : 재사용 가능, Input-> output  - 함수를 정의(def) : parameter(매개변수)  - input parameter : 함수의 입력으로 전달된 값을 받는 변수  - output parameter : return 값, 변수에 값을 다시 할당 할 수 있다.  - Arguments(인수) : 함수를 호출할 때 전달하는 입력값  - variable: 값을 담는 객체  -인수를 통해 함수값을 넘겨줄 때 순서가 중요함  05-03. 비교 연산자 : 비교 연산자(==)와 대입연산자(=) 구분해서 사용하기  05-04. 조건문 : if, elif, else  05-05. 논리 연산자 (and, or)  05-06. 반복문 :  - For in 반복범위에 있는 원소를 반복자 변수를 이용해 하나씩 순차적으로추출  - For coninue, break구문 이해하기  06\_02\_ Hisplot & Displot  -HIS: sns.histplot(data = penguins,x='flipper\_length\_mm',hue="species",multiple ='stack')  #색깔이 겹쳐서 나온다.-> 분리 옵션 multiple ='stack'  # hue option : 그룹별로 그래프에 색깔을 부여  -DIS: sns.displot(data = penguins, x = "flipper\_length\_mm",hue="species", kind="kde")  -subplot type  06\_03\_ Barplot & Countplot  Bar: sns.barplot(data=penguins, x="species", y = "flipper\_length\_mm", hue="species")  Counter: sns.countplot(data=penguins, x='species', hue ="sex")  - 그룹별 카운트를 보여주는 방법  - 그룹별로 분류했을 때 각 그룹에서 하위로 분류되는 정보(hue) 차이를 색깔을 달리해서 표시  06\_04\_ Boxplot & Violinplot : 전체적인 분포 확인, Box(25~75% 박스안), Violin(전체)  BOX: 사분위수 표시, 전체적인 분포 확인, 이상값(outlier)를 찾기 쉬움  sns.boxplot(data=penguins, x ='body\_mass\_g', y= 'species', hue='sex' )  ViolinPlot: 평균 근처에 데이터가 얼마나 있는지 전체적으로 어떻게 퍼져 있는지 모두 확인 가능  sns.violinplot(data=penguins, y ='flipper\_length\_mm', x= 'species', hue='sex' )  06\_05\_ Lineplot & Pointplot  Lineplot: 전체적인 분포 확인, x,y의 경향성 중점, 선그래프  sns.lineplot(data = penguins, y="body\_mass\_g",x ="bill\_depth\_mm", hue="species")Pointplot: 데이터와 error bar를 한번에 찍어줌, 성별의 차이 확인  sns.pointplot(data = penguins, y = "bill\_length\_mm", x ="sex", hue = "species")  06\_06\_ Scatterplot & Pairplot  Scatter: Lineplot과 비슷, 전체적인 분포 확인,데이터가 퍼져 있는 모양 중점  sns.pairplot(data=penguins, hue="species")  Pairplot:주어진 데이터의 각 feature들 사이의 관계를 표시하는 plot  sns.pairplot(data=penguins, hue="species")  06\_07\_Heatmap  - feature간 상관관계 파악, 정보차이를 색의 차이로 표시  -상관계수 메트릭스 만들기 -> Heat map 그리기 (양/음 상관관계파악)  corr = penguins.corr(),sns.heatmap(corr) |
| **퀴즈 제출** | Q1. 다음 출력값으로 다른 하나는?  1) def my\_func():  name = 'son'  print(name)  my\_func()  2) def my\_func1(name2):  name2 = 'son'  return name2  name1 ='kim'  name1 = my\_func1(name2)  print(name1)  3) def my\_func2(name) :  print(name)  name = ‘kim’  name=my\_func2(name)  print(name)  4) def my\_func3(name):  name='son'  return name  name = 'kim'  name =my\_func3(name)  print(name)  정답 : 3  - 1,2,4번 출력값은 ‘son’ 매개변수(parameter)를 return하면 전역화되어 함수 밖에서 사용할 수 있으며 일반 변수(variable)와는 구분된다.  - 3번 return이 있는 경우는 변수에 값을 다시 할당 할 수 있지만 return이 없는 경우에는 return이 없으므로 할당 할 수 없다. return값이 없는 함수를 호출한 출력값은 ‘none’이며 이것을 변수에 넣으면 none이 출력된다. ‘kim’ 과 ‘none’이 출력됨  Q2. Penguins dataframe를 전체 features들의 상관관계를 color bar을 이용하여 시각화 한 것은 ?  1) sns.histplot(data = penguins,x='flipper\_length\_mm',hue="species")  2) corr = penguins.corr()  sns.heatmap(corr)  3) sns.pairplot(data=penguins, hue="sex")  4) sns.pointplot(data = penguins, y = "flipper\_length\_mm", x ="sex", hue = "species")  정답 : 2  Heatmap은 상관 관계를 색의 차이를 이용해 표시하여 각 feature간의 상관관계를 파악하기 용이한 시각화 방법이다.  Q3. Penguins dataframe에서 종별 성별을 분리하여 물갈퀴의 길이(flipeer length) 데이터 분포를 하여 분포가 어떻게 퍼져 있는지 시각화 한 것은?  1) sns.histplot(data = penguins,x='flipper\_length\_mm',hue="species")  2) sns.boxplot(data=penguins, x ='flipper\_length\_mm', y= 'species', hue='species' )  3) sns.boxplot(data=penguins, x ='body\_mass\_g', y= 'species', hue='species' )  4) sns.violinplot(data=penguins, y ='flipper\_length\_mm', x= 'species', hue='sex' )  정답 : 4  Violonplot은 데이터의 분포의 모양을 확인하기 용이한 tool로 hue옵션에 성별을 입력하면 종별 범주가 성/별 범주로 세분화되어 시각화된다. |
| **.** |  |
| **화면 캡쳐**  **(교육 화면 + 교육일시)** |  |