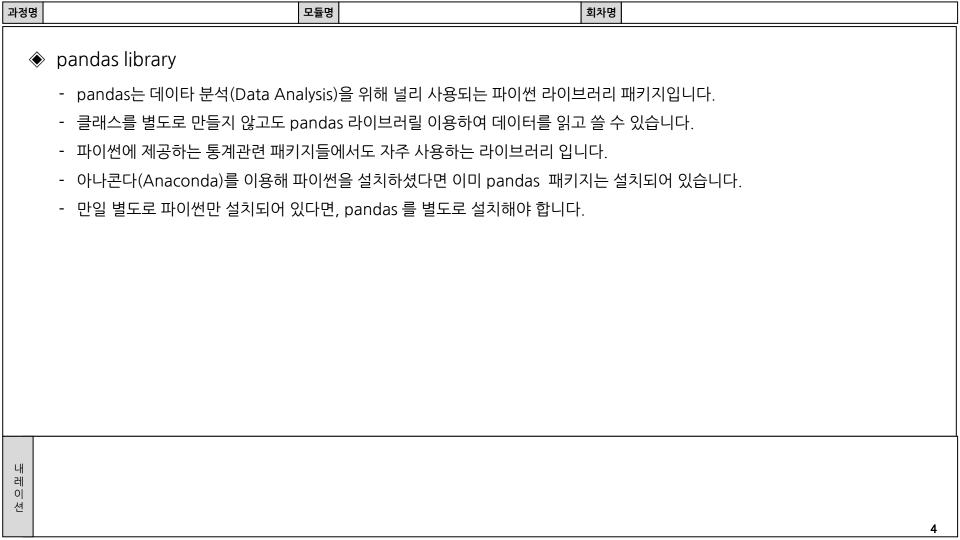
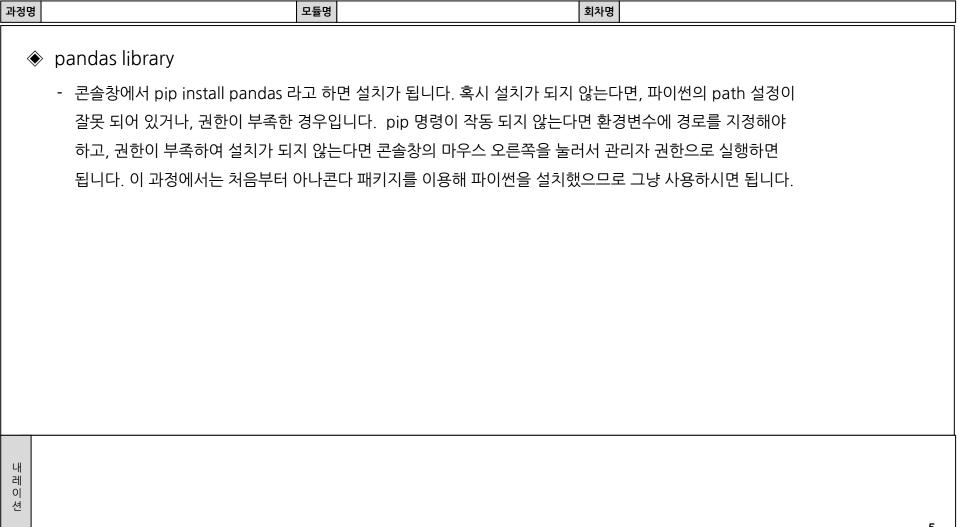
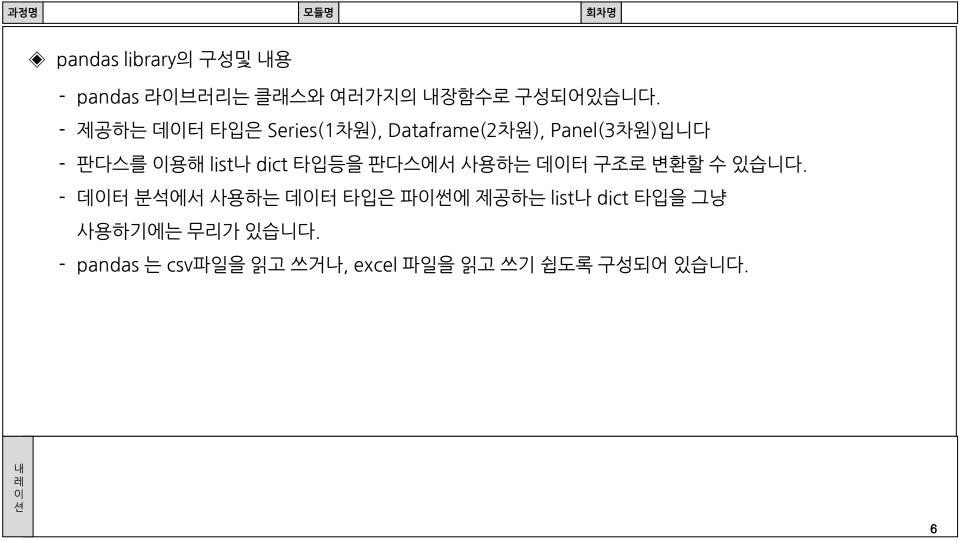
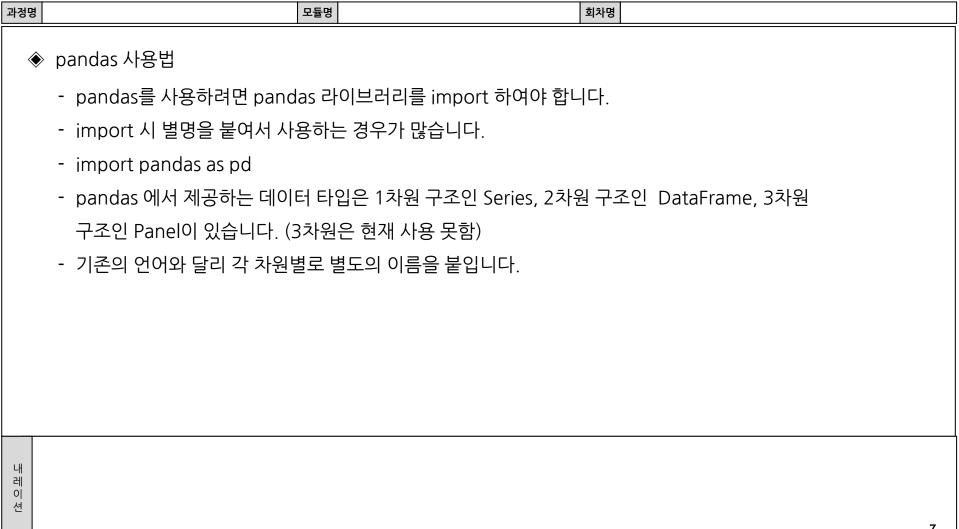


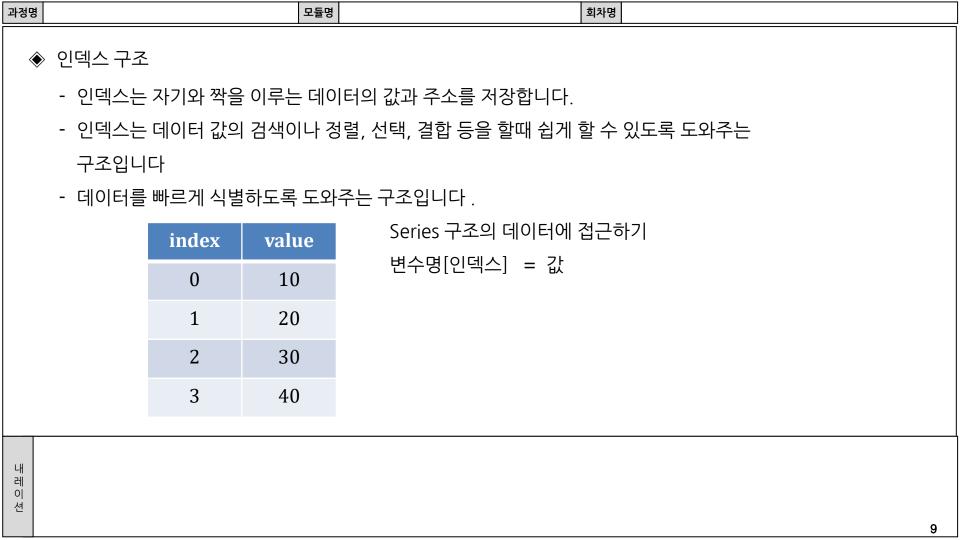
과정명	모듈명	회차명					
«	◈ 데이터셋						
	앞선 과정에서 우린 다양한 소스들로부터 데이터를 수집하는수집된 데이터들은 각자 동일하거나 다른 형태의 데이터들로						
	 데이터셋은 다향한 형태의 데이터들을 하나의 구조로 만들어서 관리나 분석을 용이하게 할 수 있습니다. 						
	 파이썬에서는 데이터셋을 관리하기 위해 pandas라는 라이브러리를 제공합니다. 데이터 분석을 위해서는 파이썬에서 기존에 사용하던 타입인 list, dict 과 같은 것 들을 그냥 						
	사용할 수는 없습니다. - 머신러닝이나 딥러닝에서는 numpy나 pandas 라이브러를 여	이용해서 파이썬이 제공하는 기본					
	타입들을 데이터 분석에 사용하기 쉽도록 제공하고 있습니다						
내 레 이 션							
		3					







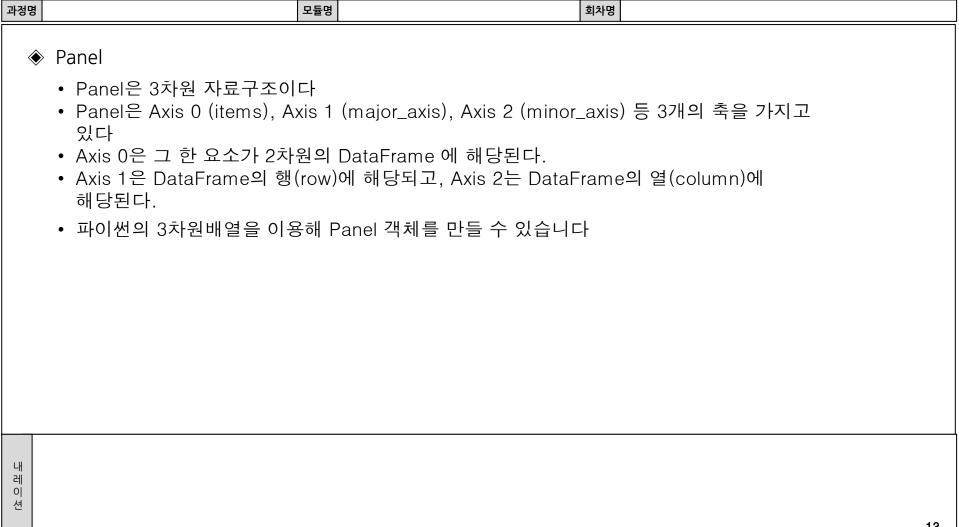




과정명	모듈명	회차명
•	Series 예제 파일명: exam9_1.py import pandas as pd #pandas 라이브러리를 사용하기 위해 import 를 한다. i #import 시 pd 라는 별도의 이름을 부여하였다. #파이썬의 list를 사용하여 Series 만들기 # 자동으로 인덱스가 부여된다 data = [10,20,30,40,50] series = pd.Series(data) print(type(series)) print(series)	<pre><class 'pandas.core.series.series'=""> 0 10 1 20 2 30 3 40 4 50 dtype: int64 a 1 b 2 c 3 d 4 e 5</class></pre>
	#직접 index 를 부여해보기 (dict타입 사용) data2 = {'a':1, 'b':2, 'c':3, 'd':4, 'e':5} series2 = pd.Series(data2) print(series2)	dtype: int64
내 레 이 션	-	10

과정명	모듈명	회차명
«	DataFrame	
	- pandas가 제공하는 2차원 구조의 타입입니다	
	- DataFrame 은 데이터를 행과 열로 이루어진 테이블의 형태로 저	장합니다.
	- pandas 의 DataFrame은 R이라는 통계언어에서 비롯되었습니	다. 다른 언어의 기능이
	파이썬으로 구현되어서 R언어보다는 기능이 좀 떨어지지만, 데0	터를 다루기 충분한 요소 들을
	가지고 있습니다.	
	- DataFrame은 열과 행으로 구성되는데 열은 각자 다른 타입으로	구성될 수 있습니다.
	- 데이터를 저장할때, 전통적인 방법이 테이블 형태로 데이터를 저	장하는 것입니다. 관계형
	데이터베이스라던가 Excel 등의 자료 그리고 csv형식의 자료들0	테이블의 구조형태입니다.
	- DataFrame은 이러한 유형의 데이터를 다루는데 최적화된 도구역	입니다.
내레	<u>-</u>	
데 이 션		
		11

과정명	모듈명	회차명
•	DataFrame 예제 파일명: exam9_2.py import pandas as pd data = { 'name':['홍길동', '임꺽정', '장길산', '홍경래'],	dict 타입의 데이터를 정의한다 dict타입은 키와 값 쌍으로 이루어진다. 특정키값에 데이터를 여러개 넣을 경우에 키:리스트 형태로 저장할 수 있다
	'kor':[90, 80, 70, 70], 'eng':[99, 98, 97, 46], 'mat':[90, 70, 70, 60], } df = pd.DataFrame(data) print("타입: ", type(df)) print(df)	타입: <class 'pandas.core.frame.dataframe'=""> name kor eng mat 0 홍길동 90 99 90 1 임꺽정 80 98 70 2 장길산 70 97 70 3 홍경래 70 46 60</class>
내 레 이 션		12



과정명		모듈명	회차명	
	 ◆ Panel 예 파일명: exam9_3.py import pandas as pd #3차원 배열을 만든다 list of list o flist [[[],[],[].]. array = [[[1,2,3], [4,5,6]], [[7,8,9], [10,11,12]], [[13,14,15], [16,17,18]]] array= pd.Panel(array) 		실행하면 경고 메시지가 나옵니다. 앞으로는 Panel 함수가 사라진다는 의미입니다 타입: <class 'pandas.core.panel.panel'=""></class>	
			<pre><class 'pandas.core.panel.panel'=""> Dimensions: 3 (items) x 2 (major_axis) x 3 (minor_axis) Items axis: 0 to 2 Major_axis axis: 0 to 1</class></pre>	
내 레 이 션			14	

모듈명 회차명	
> Series 네이터 엑세스 망법	
- Series 타입의 데이터를 접근 하는 방법은 인덱스를 활용합니다. 각각의 요소에 대한 접근 방법은 변수명[인덱스] 입니다	
- 예)	
- print(mydata1[0]) 값을 읽어서 출력하기	
- mydata1[0]=100 해당 인덱스의 값을 변경하기	
- 스크이시오 이유권 초려 하 시 ㄷ 이스니다.	
- print(myddiai[5.]) 3인당 이후도 마시락까지 물릭됩니다.	
- 인덱스가 숫자가 아니고 lable이 있을 경우 label 을 이용해 접근할 수 있습니다	
- 예)	
- print (mydata2['one'])	
	15
	Series 데이터 엑세스 방법 - Series 타입의 데이터를 접근 하는 방법은 인덱스를 활용합니다. 각각의 요소에 대한 접근 방법은 변수명[인덱스] 입니다 - 예) - print(mydata1[0]) 값을 읽어서 출력하기 - mydata1[0]=100 해당 인덱스의 값을 변경하기 - 슬라이싱을 이용해 출력 할 수 도 있습니다 - 예) - print(mydata1[0:3]) 0번방부터 3번방 이전까지 마지막 방에 있는 데이터는 제외됩니다. - print(mydata1[3]) 3번방 이전까지 출력됩니다 - print(mydata1[3:]) 3번방 이후로 마지막까지 출력됩니다. - 인덱스가 숫자가 아니고 lable이 있을 경우 label 을 이용해 접근할 수 있습니다 - 예)

과정명		모듈명	회차명
	Series 데이터 엑세스 파일명: exam9_4.py import pandas as pd data = [10,20,30,40,50,60] s = pd.Series(data) print(s[0]) #0번째 해당 데이터 출력 s[1]=200 #1번째 데이터 200으로 print(s) print(s[:5]) #5번째 데이터부터 마/print(s[2:5]) #2번데이터부터 4번 print(s[3:]) #3번 이후로 출력 data = {'one':'일', 'two':'이', 'thr series = pd.Series(data) print(series) print(series['one']) print(series['one':'three']) print(series['one':'three']) print(series['two':])	수정 시막까지 출력 데이터까지 출력	one 일 two 이 three 삼 four 사 five 오 dtype: object
내 레 이 션			16

과정명				모듈명		회차명	
•	DataFrame 데이터 엑세스 방법						
	- DataFrame은 pandas 라이브러리에서 가장 빈번하게 사용되는 데이터 타입입니다.						
	- 행과 열로 구성되어 있고, iloc, loc 라는 두개의 함수를 이용해 데이터에 접근할 수 있습니다.						
	- iloc 함	·수는 행과	열의 인덱스를	를 이용하	네 접근합니다	가. loc 함수는 행의 인덱스와 열의 컬럼명(필드명)을 통해	
	데이터	를 접근합니	니다.				
	예) 데이터	<mark>1</mark> 가 다음괴	- 같을 때				
	inde name k		kor	eng	print(mydata.iloc[0, 0])		
		X			print(mydata.loc[0, 'name'])	print(mydata.loc[0, 'name'])	
		0	홍길동	90	85	print(mydata.iloc[1, 2])	
		1	장길산	80	75	print(mydata.loc[1, 'kor'])	
		2	임꺽정	70	65		
		3	홍경래	60	55		
내							
레이							
션							17

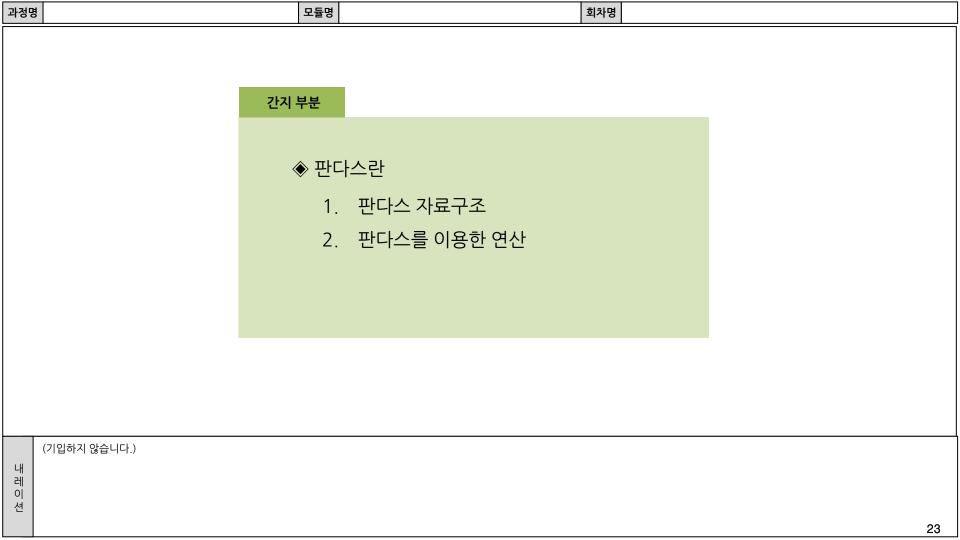
과정명	모듈명 회차명	
•	DataFrame 데이터 엑세스 예제 파일명 : exam9_5.py	
	import pandas as pd data = { ' name ' :[' 홍길동 ' , ' 임꺽정 ' , ' 장길산 ' , ' 홍경래 ' , ' 이상민 ' , ' 김수경 '], ' kor ' :[90, 80, 70, 70, 60, 70], ' eng ' :[99, 98, 97, 46, 77, 56], ' mat ' :[90, 70, 70, 60, 88, 99], } df = pd.DataFrame(data)	
	print(df.head()) #앞의 다섯명에 대한 데이터만 나온다.	
	print(" 지정한 열만 출력 ") print(df[' name ']) print(df[' kor ']) print(df.columns) #컬럼이름만 출력이어서	
내 레 이 션		18

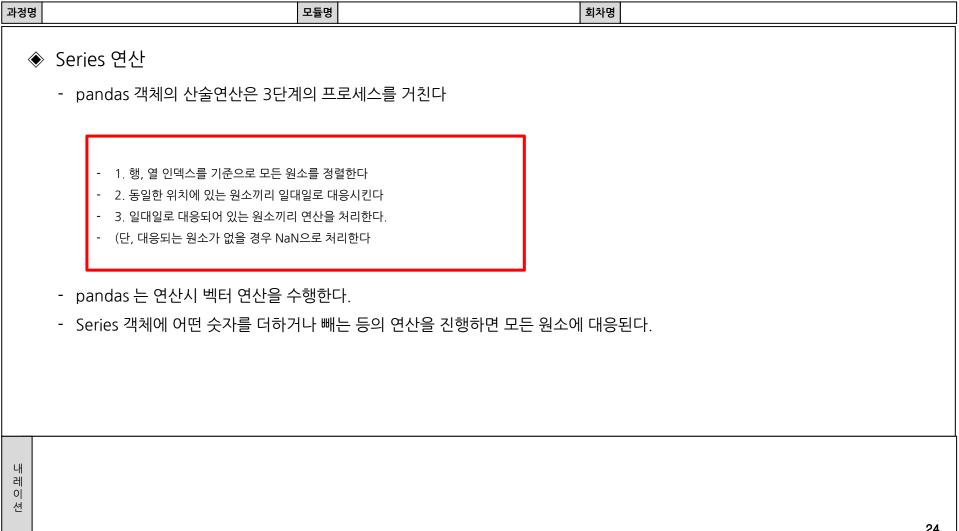
과정명	5	고듈명	회차명
•	→ DataFrame 데이터 엑세스 예제 파일명 : exam9_5.py		
	#iloc 함수: 배열에서의 위치값으로 (print("iloc 함수 사용") print(df.iloc[0,0]) #0,0번에 해당하 print(df.iloc[3,2]) #3번째 행의 2번 print(df.iloc[2:4,2]) #2번행의 2번	는 데이터 출력하기 째 열	
	print(df.iloc[2:4,2:4]) #2번행의 2,	3번열, 3번행의 2,3번열	
	print("loc함수 사용") #loc 함수는 필드명으로 데이터를 출력 print(df.loc[0, 'name']) #0번째 행 print(df.loc[3, 'eng']) #0번째 행의 print(df.loc[:, 'name':'eng']) #슬리	의 name필드 값 출력	출력하라
내 레 이 션			19

과정명	모듈명		회차명			
•	- Panel은 3개의 축을 가지고 있다					
	 , Panel 의 데이터 구조를 Dict 타입으로 다음처럼 나타낼 data = { 'class1':{ 		lass2			
	'name': ['박동석', '권다윤', '김민제'], 'height':[173, 177, 183]	class1 i	ndex	name	height	
	}, 'class2'∶{	index	name	height	9	
	'name': ['김민수', '이용희', '임미영'], 'height':[179, 167, 180]	0	박동 석	173	7	
	} }	1	권다윤	177	0	
	panel = pd.Panel(data)	2	김민제	183		
	class1 = panel['class1']					
	print(class1.iloc[0,0])					
내 레 이 션						

과정명	모듈명	회차명
*	Panel 예제 파일명 : exam9_6.py import pandas as pd	
	data = {	
	panel = pd.Panel(data) print(panel) 이어서	
내 레 이 션		21

과정명		모듈명 회차명	
	Panel 예세 파일명:exam9_6.py df = panel['class1'] print(df.iloc[0,0]) print(" for") for key in panel: item = panel[key] for col in item.columns: for row in item[col]: print(row, end=' ') print()	Dimensions: 2 (items) x 3 (maj Items axis: class1 to class2 Major_axis axis: 0 to 2 Minor_axis axis: name to heigh name height 0 박동석 173 1 권다윤 177 2 김민제 183 name height 0 김민수 179 1 이용희 167 2 임미영 180 박동석 Index(['name', 'height'], dtyp for 박동석 권다윤 김민제 173 177 183 김민수 이용희 임미영 179 167 180	t
내 레 이 션			22





과정명		모듈명	회차명
•	Series 연산 1 파일염: exam9_7.py import pandas as pd data1 = {'kor':90, 'eng':70, 'mat':80} data2 = {'kor':90, 'eng':70, 'mat':80} data3 = {'kor':90, 'eng':70, 'mat':80} data4 = {'eng':90, 'mat':70, 'kor':80} series1 = pd.Series(data1) series2 = pd.Series(data2) series3 = pd.Series(data3) series4 = pd.Series(data4) result1 = series1 + series2 + series3 + series result2 = result1/4 print("총점") print(result1) print("평균") print(result2)	#인덱스의 순서가 바뀌어도 정렬을 진행하: s4	기때문에 알아서 대응된다. 총점 eng 300 kor 350 mat 310 dtype: int64 평균 eng 75.0 kor 87.5 mat 77.5 dtype: float64
내 레 이 션			25

과정명	_	모듈명	회차명
	Series 연산 2 파일명: exam9_8.py import pandas as pd #인덱스가 없을 경우 값이 NaN으로 들어가서 data1 = {'kor':90, 'mat':80} data2 = {'kor':90, 'eng':70} data3 = {'kor':90, 'eng':70, 'mat':80} series1 = pd.Series(data1) series2 = pd.Series(data2) series3 = pd.Series(data3) result1 = series1 + series2 + series3 print(result1)	연산이 되지 않는다 eng N kor 270	laN J. Ø laN
내레이션			26

과정명		모듈명	회차명
	Series 연산 3 파일명: exam9_9.py import pandas as pd #인덱스가 없을 경우 값이 NaN으로 들어가서 #add 연산시 fill_value=0 옵션을 주면 대응되- #지정한 값으로 대신해준다 data1 = {'kor':90, 'mat':80} data2 = {'kor':90, 'eng':70} data3 = {'kor':90, 'eng':70, 'mat':80} series1 = pd.Series(data1) series2 = pd.Series(data2) series3 = pd.Series(data3) result1 = series1.add(series2, fill_value=0).a	는 셀에 값이 없을 경우 fill_value에서	eng 140.0 kor 270.0 mat 160.0 dtype: float64
내 레 이 션			27

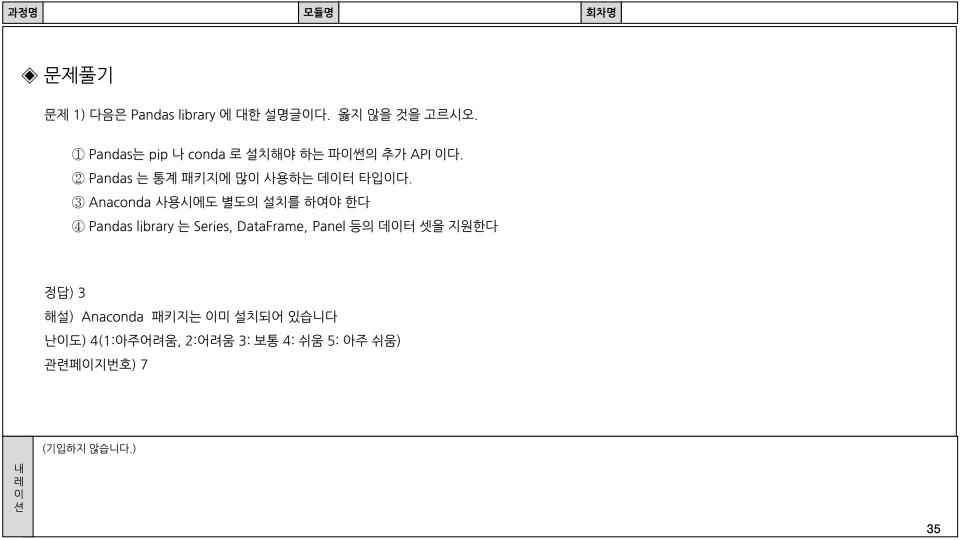
과정명	모듈명 회차명						
•	 ◆ DataFrame 연산 - DataFrame은 여러개의 Series의 결합으로 볼 수 있습니다. 그래서 연산 방법도 Series의 내용이 적용됩니다. - 산술연산이 모두 가능합니다. 						
	- 1. 행, 열 인덱스를 기준으로 모든 원소를 정렬한다 - 2. 동일한 위치에 있는 원소끼리 일대일로 대응시킨다 - 3. 일대일로 대응되어 있는 원소끼리 연산을 처리한다 (단, 대응되는 원소가 없을 경우 NaN으로 처리한다						
	 새로운 행을 추가하거나 삭제가 가능합니다. 주의사항은 새로운 행을 추가하거나 삭제시 결과 값을 반환받아야 합니다. 결과값을 반환받지 않으면 추가되거나 삭제된 행은 원래대로 보입니다 데이터 프레임에 새로운 필드를 손쉽게 붙일 수 있습니다 						
내 레 이 션		28					

과정명		모듈명	회차명
	DataFrame 데이터 엑세스 예제 파일명: exam9_11.py # dict 타입을데이터 프레임으로 전환후 새로	² 운 필드를 추가하기	
	파일명: exam9_6.py import pandas as pd data = {	'홍경래', '이상민', '김수경'],	
내 레 이 션			30

과정명		모듈명		회차명	
•	DataFrame 연산에게 파일명: exam9_12.py import pandas as pd data = { 'fruits':['망고', '딸기', '수박', '파인애플'], 'price':[2500, 5000,10000, 7000], 'count':[5, 2, 2, 4], } df = pd.DataFrame(data) #새로운 행 추가하기, 주의: 반드시 반환되는 #append 함수는 데이터가 추가된 객체를 반함 df.append({'fruits':'사과', 'price':3500, 'couprint(df) #ignore_index=True 옵션: 기존의 인덱스를 - print("데이터가 추가된 객체를 받을 경우 df = df.append({'fruits':'사과', 'price':'3500 print(df)	환한다 . int':10] 무시하리 	· , ignore_index=True) 라는 옵션인데 붙여야 한다 ')		
내 레 이 션					

과정명		모듈명	회차명
•	2 3 (3) 1 3 11 1 2		fruits price count 0 망고 2500 5 1 딸기 5000 2 2 수박 10000 2 3 파인애플 7000 4 데이터가 추가된 객체를 받을 경우 fruits price count 0 망고 2500 5 1 딸기 5000 2 2 수박 10000 2 3 파인애플 7000 4 4 사과 3500 10 fruits count 0 망고 5 1 딸기 2 2 수박 2 3 파인애플 4 4 사과 10 원본 fruits price count 0 망고 2500 5 1 딸기 5000 2 2 수박 10000 2 3 파인애플 4 4 사과 10 원본 fruits price count 0 망고 2500 5 1 딸기 5000 2 2 수박 10000 2 3 파인애플 7000 4 4 사과 3500 10 fruits price count 1 딸기 5000 2 2 수박 10000 2 3 파인애플 7000 4 4 사과 3500 10 fruits price count 1 딸기 5000 2 2 수박 10000 2 3 파인애플 7000 4 4 사과 3500 10
내 레 이 션			32

			모듈명	회차명			
요하기							
공이기							
다음 테이블 데이터를 데이터프레임 객체로 만들어 봅시다.							
X2.	Х3	X4					
			2				
			3				
2	7.2	10.8	4				
			2				
3.2	9.6	12.6	3				
이블에 마지막 요	소로 다음	데이터를 추기	하세요				
(10, 20, 30, 40)							
위 테이블에 total 필드를 만들고 각 필드들의 합을 구하여 total							
필드에 적용하세요							
지 않습니다.)							
					33		
	X2 2.9 2.4 2 2.3 3.2 기블에 마지막 요 20, 30, 40) 기블에 total 필드	N2 X3 2.9 9.2 2.4 8.7 2 7.2 2.3 8.5 3.2 9.6 N	M이블 데이터를 데이터프레임 객체로 만 X2 X3 X4 2.9 9.2 13.2 2.4 8.7 11.5 2 7.2 10.8 2.3 8.5 12.3 3.2 9.6 12.6 이블에 마지막 요소로 다음 데이터를 추기 20, 30, 40) 이블에 total 필드를 만들고 각 필드들의 현 작용하세요	## ## ## ## ## ## ## ## ## ## ## ## ##	## Note ## No		



번호	다음에 제시된 파서중 추가로 설치하지 않아도 사용 가능한 기본 파서는?	정답	난이도	해설	관련학습보기
2	다음과 같이 Serirs 객체가 있을 경우 데이터 출력을 하려면 어떻게 해야야 하나? series = pd.Series(['사과', '망고', '바나나', '수박', '딸기']) print () 결과: 바나나 수박 딸기	series[2:	4	Series 객체도 슬라이싱이 적용됩니니다 2,3,4 세개의 데이터가 출력되어야 하므로 2번방부터 마지막까 지 적용됩니다.	32
3	다음과 같이 Serirs 객체가 있을 경우 데이터 출력을 하려면 어떻게 해야야 하나? series2 = pd.Series(('red':'빨간색', 'green':'초록색', 'blue':'파란색', 'black':'검은색')) print () 빨간색, 초록색, 파란색	series2['r ed':'blue']	4	인덱스가 문자열이므로 문자열로 주면 된다. series2['red':'blue']	29
4	data = { 'name':['Tom', 'Brown', 'Jane', 'Stepany', 'John'], 'month1':[1200, 2000, 3000, 1800, 4500], 'month2':[2400, 2300, 4300, 2800, 4200], 'month3':[1800, 2700, 3500, 3800, 4300], } data = pd.DataFrame(data) 위의 데이터는 ¼ 분이 영업자들의 실적이다. 이 실적을 새로운 필드를 추가하여 평균을 구하는 코드를 작성하시오	정답은 해설참조	4	data['quater1'] = (data['month1'] + data['month2'] + data['month3'])/3 또는 data['quater1'] = (data.month1+ data.month2+ data.month3)/3	6
5	위 6번의 데이터에 새로운 사람의 자료를 추가하는 코드를 작성하세요 name='Bread', month1=3400, month2 = 4000, month3=5000	정답은 해설참조	5	data = data.append({'name':'Bread', 'month1':3400, 'month2':4000, 'month3':5000}, ignore_index=True)	32

퓝여	문제	정답	난이 도	해설	관련학습보기
6	위 4번의 데이터중에서 Stepany 의 데이터중 3월 데이터를 보고 싶다면 어떻게 해야 하는지 코드를 작성하시오(iloc 함수 사용하기)	정답은 해설참조	4	data.iloc[3, 3]	31
7	위 4번의 데이터중에서 john 의 데이터중 2월 데이터를 보고 싶다면 어떻게 해야 하는지 코드를 작성하시오(loc 함수 사용하기)	정답은 해설참조	4	data.loc[4, 'month2']	33
8	4번의 데이터에서 , month3 필드를 삭제하는 코드를 작성하시오	정답은 해설참조	4	data = data.drop('month2', axis=1)	28
9	4번의 데이터에서 jane 의 데이터를 삭제하는 코드를 작성하시오		5	data = data.drop(2, axis=0)	34
10	다음 설명중 잘못 된것은?. ① Series 객체는 list를 통해서만 만들 수 있다 ② Series 객체는 dict 타입을 통해 직접 index 를 부여할 수 있다 ③ DataFrame은 행과 열로 구성된 테이블 형태 구조체이다 ④ Panel 은 3차원 데이터 구조이다	1	2	list, dict 타입 둘다 가능합니다	30

