

실습05 데이터 정규식 처리

교육 안내

- 접속주소 : <https://rpakr.com/lab/datatable2>
 - 실습순서안내
1. 웹 주소에서 표 데이터 스크래핑 : 지정된 웹 주소로부터 표 형태의 데이터를 추출합니다.
 2. 데이터 유효성 확인: 아래는 유효한 데이터 조건입니다. 해당 조건에 모두 부합하는지 확인하세요.
 - 번호%이름 :
번호와 이름이 특수문자로 합쳐져 있습니다.
이름은 2~6 글자수의 한글로만 작성되어 있습니다.
 - 생년월일 :
'yyyy-MM-dd' 형식입니다.
작업 수행일 기준 만 19세이상입니다.
만약 유효한 생년월일이 아니며 주민등록번호의 앞자리의 생년월일이 유효할 시,
생년월일을 주민등록번호 앞자리로 변경합니다.
 - 주민등록번호 :
xxxxxx-xxxxxxx 형식입니다.
앞번호는 생년월일과 일치해야 하며, 뒷 번호의 첫 번째 숫자는 1~4 값입니다.
만약, 주민등록번호의 앞자리가 생년월일과 동일하지 않을 시, 생년월일로 주민등록번호의 앞자리를 변경합니다.
 - 전화번호 : '000-0000-0000' 형식입니다.
 - 이메일 : xxx@xxx.xxx 형식입니다.
 - 응시지역 :
도로명주소 홈페이지(<https://www.juso.go.kr/openIndexPage.do>)에서 조회 후 검색결과가 존재해야 합니다.
 - 응시번호 :
8자리로 대문자 알파벳과 숫자로만 이루어집니다.
대문자 알파벳과 숫자가 각각 한 번 이상 포함되어 있는지 확인합니다.
 - 국,영,수, 금액
숫자로만 이루어집니다.
 3. 검증 열 추가: 검증 열을 추가하여 행의 모든 데이터가 유효할 경우 'O', 아닐 경우 사유를 작성합니다.
 - * 사유가 2개 이상일 경우, 모두 작성합니다.
 - * 만약 생년월일 혹은 주민등록번호에 변경이 있을 경우, 변경 내용 또한 작성합니다.
 4. 결과 파일 생성 : 조건에 맞게 최종 결과 파일을 작성합니다.
 - 4-1) 번호와 이름을 나누어 각각 다른 열에 저장합니다.
 - 4-2) 평균 열을 추가하여 국어, 영어, 수학의 평균 점수를 도출합니다.
 - 4-3) 평균 점수가 50점 미만이라면 빨간색을, 이상이라면 초록색을 평균 열에 한정하여 지정합니다.
 - 4-4) 금액 열에 원화표시와 콤마 표시를 넣습니다.
 - 4-5) '오류' 시트를 생성하여 유효하지 않은 데이터만 작성합니다.
 - 4-6) 바탕화면에 "데이터테이블_YYYYMMDD.xlsx"으로 저장합니다.

실습05 데이터테이블 정규식 처리

번호	번호이름	생년월일	주민등록번호	전화번호	이메일
1%	김수민	1943-04-14			
2%	이도현	2023-01-14	230114-392688	020-1012-341111	user746
3%	박민서	1934-04-22	340422-1580545	020-1111-2222	.awfe@dafx.co
4%	최예준	1978-02-27	780227-1036459	020-111-1-111-2222	user8720@gm.
5%	장하린	1932-08-26	320826-1849160	020-111-22221111	kim@naber.cor
6%	윤지우	1998-12-12	981212-1264556	020-1234-5678	user3854@gm
7%	정다은	1997-09-23	970923-1315517	020-010-5-122-5679	user848@dma
8%	강	1916-10-14	161014-1355312	020-2221-5681AA	user7728@yah
9%	오현우	1968-11-14	681114-1399619	020-0521-5681	user995@nabe
10%	서지호	1918-03-11	180311-1718666	020--723-5682	user6563@dm
11%	준박	1907-03-29	070329-1595180	020-1234-568	user4319@gm:

현재 페이지: 1page / 전체 페이지: 5page

모든 표 데이터를 스크래핑

조건에 맞춰서 데이터테이블 가공

열의 데이터 유효성 검사

생년월일과 주민등록번호 일치

결과 및 오류 내용 작성

번호	번호이름	생년월일	주민등록번호	전화번호	이메일	국어	영어	수학	응시지역	응시번호	금액	평균	검증 결과	오류 사유
1	김수민	1943-04-14	430414-1136266	1012341111	user2393@example.com	11	75	50	대전광역시 서구 둔산로 100	A123D21Q	₩49,800	45.33333333	FALSE	* 전화번호가 유효한 형식이 아닙니다.
2	이도현	2023-01-14	230114-392688	020-1012-341111	user746	2	41	4	서울특별시 기리구 오리로21길	13829933	₩7,300	15.66666667	FALSE	* 만 19세 이상이 아닙니다. * 주민등록번호 형식이 아닙니다. * 전화번호가 유효한 형식이 아닙니다. * 올바른 이메일 형식이 아닙니다. * 유효한 주소가 아닙니다.
3	박민서	1934-04-22	340422-1580545	020-1111-2222	awfe@dafx.com	87	16	20	경기도 수원시 권양구 금곡로 127	A3aa1122	₩29,500	41	FALSE	* 유효한 주소가 아닙니다. * 응시번호에 숫자와 영어 대문자 외 문자가 포함되어 있습니다.
4	최예준	1978-02-27	780227-1036459	020-111-1-111-2222	user8720@gmail.com	23	57	50	경기도 성남시 중원구 성남대로 997	BBA1K3OW	₩10,600	43.33333333	FALSE	* 전화번호가 유효한 형식이 아닙니다.
5	장하린	1932-08-26	320826-1849160	020-111-22221111	kim@naber.com	44	71	81	경기도 아산시 아마읍 양평중학교길 29	a133AAQQ	₩30,200	65.33333333	FALSE	* 전화번호가 유효한 형식이 아닙니다. * 유효한 주소가 아닙니다. * 응시번호에 숫자와 영어 대문자 외 문자가 포함되어 있습니다.
6	윤지우	1998-12-12	981212-1264556	020-1234-5678	user3854@gmail.com	56	81	0	울산광역시 중구 중가5길 30	BAQZ22SW	₩16,000	45.66666667	TRUE	
7	정다은	1997-09-23	970923-1315517	020-010-5-122-5679	user848@dmil.net	32	60	7	경기도 안양시 동안구 임곡로80번길 34	0O9IUUJW	₩18,500	38	FALSE	* 전화번호가 유효한 형식이 아닙니다.
8	강	1916-10-14	161014-1355312	020-2221-5681AA										

..\바탕화면\ 데이터테이블_YYYYMMDD.xlsx 저장