

[양식2] 근태기록부

2019 공공 빅데이터 청년 인턴십

일일근태기록부

(2019년 11월)

참여기관: 화성시 수련생: 김동혁 기관담당자 : 이지호

일자	요일	성명	출근시간	퇴근시간	근태사항					확인 (수련생)	담당자 확인
					결근	지각	조퇴	외근	휴가		
16	토	—	—	—							
17	일	—	—	—							
18	월	김동혁	9:00	18:00						(pm) 임지호	
19	화	김동혁	9:00	18:00						(pm) 임지호	
20	수	김동혁	9:00	18:00						(pm) 임지호	
21	목	김동혁	9:00	18:00						(pm) 임지호	
22	금	김동혁	9:00	18:00						(pm) 임지호	
23	토	—	—	—							
24	일	—	—	—							
25	월	김동혁	9:00	18:00						(pm) 임지호	
26	화	김동혁	9:00	18:00						(pm) 임지호	
27	수	김동혁	9:00	18:00						(pm) 임지호	
28	목	김동혁	9:00	18:00						(pm) 임지호	
29	금	김동혁	9:00	18:00						(pm) 임지호	
30	토	—	—	—							
31											
합계										1 (pm) 임지호	

[수련생-서식 3] 주간·월간 활동보고서 양식

2019 공공 빅데이터 청년 인턴십
인턴십 프로젝트 주간 활동보고서
(2019년 11월)

참여기관: 화성시 수련생 성명: 김동혁

활동 일시	월-주차	기간	참여기관 담당자 확인
	11월 1주차	2019.11.1.(금)	
	1일차 (11.1)	금일에는 CCTV 초·중·고교에서 80% 확보하였다. 최종 위치순위를 로드부과 위성사진으로 간단한데 정리하면 될것이다. 이후에는 통계조사상황실에 가서 필요한 아파와 사용용품, 조사용품등을 전달해주었다. 복구하고 API를 준비하였고 필요한 정보를 파일별로 통계수집하였다. 그후 CCTV분석이 끝나면 데이터베이스 관리 및 SQL언어를 공부하여 DB 관리연습을 해보려 한다.	이지호
활동 내용			

[수련생-서식 3] 주간·월간 활동보고서 양식

2019 공공 빅데이터 청년 인턴십
인턴십 프로젝트 주간 활동보고서
(2019년 11월)

참여기관: 화성시 수련생 성명: 김동혁

활동 일시	월-주차	기간	참여기관 담당자 확인
	11월 2주차	2019.11.4.(월)~2019.11.8.(금)	
활동 내용	1일차 (11.4)	금일은 데이터 분석 분야를 공부를 하였다. 통계분석파트에서 기본통계 개념부터 측정방법, 분포 추정 및 가설검정, 회귀분석, 주성분분석, 시계열데이터 다루는 방법에 대해 학습하였다. 시계열 데이터를 한번도 다룬적이 없어서 개념학습이 어려웠지만 응용문제를 풀어봄으로서 어느 정도 내용을 머릿속에 적립하였다. 데이터분석대학원에 있어서 차원축소, 변수선택법이 매우 중요한 요소인데, 여러가지 방법이 있다는 것을 알았다.	수련생
	2일차 (11.5)	금일오전에는 각 지역 통계상황실에 필요 물품 전달 및 조사가 시작되어 어려움이 있거나 상황을 관리하기 위해서 출장을 다녀왔다. 또한 낭부에 위치한 여러 부서에 올해 작성된 통계책자를 전달하였다. 오후에 본격하여 ADSP 학습을 하였다. 시계열 데이터, 정상성과 비정상성 시계열의 차점과 예측·분산에 따른 차이, 변환방법에 대해 공부하였다. 유동인구 같이 시계열성을 띠는 데이터에서 활용하여 인구변동을 해석도 중얼거렸다.	이지호
	3일차 (11.6)	금일은 정형데이터 마이닝에 대해 공부하였다. 지도학습과 비지도학습의 분석 방법론과 각 분석의 개념을 학습하였다. CCTV 분석을 할 때 설명 변수들을 활용하여 규칙분석을 진행하였는데 어떤 방식으로 분석이 진행되나지, 변수의 특성별로 사용해야 하는 방법이 차이가 있다는 것을 알았다. 이후 영향규칙과 측정지표(지지도, 신뢰도, 향상도 등) 계산 방법을 공부하여 해당 영역 문제를 풀어 정리를 하였다.	이지호
	4일차 (11.7)	금일은 비정형 데이터 마이닝에 대해 공부하였다. 기존에 축적되어 있던 데이터들은 대부분 정형 데이터이며 시간·지역·주제·비정형 데이터의 양이 방대해지고, 형태와 구조가 매우 복잡하여 쉽게 접근할 수 있지 않았다. 헉스로 마이닝을 알아보는데, 벤드분석·트리분석·군집분석·비슷한분류방법 등을 확습했다. 이전에 분석한 회사의 판권분석데이터를 벤드분석 정도를 해보며 기초를 다졌다.	이지호
	5일차 (11.8)	금일도 비정형 데이터 마이닝을 학습하였다. 군집분석을 좀더 구체적으로 공부하였고, 어떤 거리계도를 사용하는지 여러가지 군집 알고리즘 중 무엇을 사용할 것인지·최적의 군집수는 몇개가 적당한 것인지에 대해 알았다. 통계학 전공을 하면서 분석과 프로그래밍 수업을 주로 들었고 선형대수학, 행렬을 접한 경험이 드물어 기본적으로 데이터를 볼 때 두 가지를 사용하여 변수가 어떤 관계를 가지는지 알아야 할 것이라고 생각했다.	이지호

[수련생-서식 3] 주간·월간 활동보고서 양식

2019 공공 빅데이터 청년 인턴십
인턴십 프로젝트 주간 활동보고서
(2019년 11월)

참여기관: 화성시 수련생 성명: 김동혁

활동 일시	월-주차	기간	참여기관
			담당자 확인
활동 내용	1일차 (11.11)	금일 오전에는 제한주에 이어서 선행대수학과 행렬에 대해 공부하였다. 텁성과 풍생의 연산후에 나타나는 변화와 구조에 초점을 두는 것인데, 주어진 데이터(벡터)를 선형사상(일차원)에서 표현하는 방식이라고 한다. 흔히 코드에서 말하는 시각화와 유사한데, 가령 원점을 지나는 직선에 대한 대칭변환(reshape), 수직임사등에 기초적인 이론을 이해하고 있다. 오후에는 통계상학으로 출장을 다녀왔고, 내일 이어서 학습할 예정이다.	○) 지
	2일차 (11.12)	금일에도 선행대수를 이어서 학습했다. 3차원에서의 기저벡터를 표현하면서 차원축소를 위해 정규직교하는 과정을 학습했다. 기저벡터끼리 내적시 크기가 0이 되는것이 바로 좌표계 인데, 여태 gis를 사용하면서 공부했던 좌표계와는 다른 개념이었다. gis에서 사용하는 polynomial도 종류 6가지 를 만족한다면 벡터라고 한다. 두개의 함수를 사용해도 표현이 가능하다는점이 신기했다. 오후에는 분석가능성이 빤나워서와 분석데이터를 공유해주셨다.	○) 지
	3일차 (11.13)	금일 오전에는 통계연보 작성을 도왔다. 사법체 지표를 지역별, 연령별, 성별 등으로 구분되어 있는 자료를 피벗테이블을 활용하여 기준형식에 맞도록 정리를 하였다. 오후에는 전날 분석가능이 정리해준 데이터를 활용하여 분석보고서를 수정하였다. 그중, 범죄통계데이터 및 토지이용데이터 같은 일부에서 새로 받은 데이터가 있었고, 그 사용에 사용해야 할지 EDA 작업의 필요성을 느꼈다.	○) 지
	4일차 (11.14)	금일 오전에는 분석보고서를 수정하기 앞서 새로받은 데이터를 병합하여 EDA를 하는 단계로 병합하고 5대 병목과 발생한 빈도를 표출하여 등급화된 데이터를 명령하였다. 이어 토지사용특성에 따른 계시별 면적을 다시 병합하고, 이전에 가지고 있던 삽가입도를 첨, 범위·유치원·공원·은행·ATM·학교등 모두 차표에 찍어보며 경계별 개수를 병합하였다. 어떤 지역자, 행정자인 표와 함께 발생에 영향을 미치는지 알아볼 듯, 설치밀도와 기장치로 부여해 데한다.	○) 지
	5일차 (11.15)	휴가	○) 지

[수련생-서식 3] 주간·월간 활동보고서 양식

2019 공공 빅데이터 청년 인턴십
인턴십 프로젝트 주간 활동보고서
(2019년 11월)

참여기관: 화성시 수련생 성명: 김동혁

활동 일시	월-주차	기간	참여기관
			담당자 확인
활동 내용	11월 4주차	2019.11.18(월)~2019.11.22.(금)	
	1일차 (11.18)	금일 오전에는 지난주 분석했던 자료를 다시 정리하였고, 기준에 작성했던 보고서와는 다르게 분석가님이 정해주신 양식과 기초데이터를 다시 좌표에 찍어 그림을 고쳐해야 하는 작업이 꽤 오래걸렸다. 특히 중간에 범죄등급을 분석하는데 경자의 어떤 특성도 범죄발생등급과는 상관성이 있어서, 추가로 할 분석작업은 없을 것 같았다. 또한 <u>불교준분석모델</u> 과는 다르게 범죄데이터를 사용함으로서 근거마연에 정보를 좀더 찾아 추가해야 할 것 같다.	인 지 흐
	2일차 (11.19)	금일은 오전오후 모두 출장을 다녀왔다. 오전에는 아래와 사용품들을 준비하여, 각 통계 상황실에 들었다. 오후에는 목인일에 진행하는 공무원 대상 교육을 하는데 필요한 용품을 구입하였다. 필요 갯수가 많아서 시간이 많이 소요되었다. 사무실로 다시 보고서 작성을 마무리하였다. 현재 도로여 주제와 소외인지역자들을 범람할 반면에 대안지가 많아서 찾는 편이다. 각 지역별로 도로가 꾸준히 있다면 <u>top100 리스트</u> 를 정기할 수 있을 것 같다.	인 지 흐
	3일차 (11.20)	금일은 전화표와 주거지를 주소와 자료들에 변환할 수 있는 자료인 <u>프로그램</u> 을 찾아보았고, 굉장히 복잡한 절차로 이루어서 있어 학습하는데 시간이 걸렸다. 상에서 전화표를 경자의 중심집으로 찾았고 그정도 <u>gis</u> 가 자동으로 인식하여 위, 경도 좌표를 찍어주는 툴을 사용하였다. 이후 계간 분석보조프로그램을 사용하여 위경도를 주소로 변경하여 리스트를 만드는데 큰 도움이 되었다. 내일주로 보고서가 완성될 것 같다.	인 지 흐
	4일차 (11.21)	금일 오전에는 <u>top100 리스트</u> 를 선출한 뒤, 상위 일정의 조트별 차지도 이어지를 정리하여 보고서를 확정작성하였다. 분석가님의 피드백을 받았고, 수정할 부분은 있지만 조금 보충설명이 필요한 부분은 짚어주었다. 오후에는 빅데이터 교육준비를 조율하고 듣기로 했다. 국가공감안전원에서 원장님이 직접 특강을 하셨고, 빅데이터에 대한 호기심은 불러일으키는 유익한 특강이였다.	개 체 워크
	5일차 (11.22)	금일 오전에는 어제 원장님이 요청받은 화성시 공모사업 서각화를 드렸다. 이전에 자료를 보낸 이후에 공모사업 전수의 개수에 변동이 있었고, 그래서 자료중에 부서명이 바뀌어서 편집이 필요했다. 추가로 3번 부서별 공모사업선정별 금액도 추가로 원하셔서 세탁작업을 제출하였다. 오후에는 내일 춘천문화시행이어서 공부를 하였다. 통계분석부분은 맨먼저지만 분석기획과 거버넌스 PART는 아직 보통이 필요하다.	인 지 흐

[수련생-서식 3] 주간·월간 활동보고서 양식

2019 공공 빅데이터 청년 인턴십
인턴십 프로젝트 주간 활동보고서
(2019년 11월)

참여기관: 화성시 수련생 성명: 김동혁

활동 일시	월-주차	기간	참여기관
			담당자 확인
활동 내용	1일차 (11.25)	금일에는 분석보고서를 최종적으로 피드백을 받았다. 분석가능이 작성 코딩하여 추출한 화성시 top 100개소와 내가 뽑은 top100 개소의 위치가 다른부분이 있어서 데이터를 흥행하고, 보고서를 완부 수정해야 하는 부분이 생겼다. 혜안에 올릴 보고서로는 경찰청에서 제공하는 범죄내용 데이터를 아울러 수정을 해야한다. 기한은 29일 오후 11시로 설정으며 분석가능이 코드에 사용한 테이터를 받음과 동시에 gis, 보고서를 마무리하면 될 것 같다.	인 자 초
	2일차 (11.26)	금일은 혜안용 보고서 수정을 하였다. 경마보고용과 혜안용을 따로 작성하였으며, gis상에 보이는 범죄내용을 아울러 수정하였다. 어느정도까지가 유통이 되는지 몰라서 제일 큰 범주인 동부와 서부에서의 범죄발생 건수 정도로 마스킹하였다. 이후 금연등 정보가 들어가지 않지만 범례가 0~10등급으로 알아내기 발생했는지 반드시 그램 정도로 마무리하였다.	인 자 초
	3일차 (11.27)	금일은 K-ICT 데이터 그랜드 컨퍼런스에 다녀왔다. 탐문보도와 함께 참석하였으며, 여러 기업에서 데이터 분석업무를 설명해주는 연사들이 조끼와 있고 해당기업에서 실시하는 프로세스에 대해 강의하였다. 아래로 보발나라의 개인화성, 적합성, 정확성등의 기준을 설명해준것이 크게 와닿았다. 뿐만 아니라 지수(선호지수, 소비자수등)를 예상해내는 의미를 두는 것 같았다. 데이터를 가지고 분석하는 일에 큰 자부심을 느껴야 한다고 생각했다.	인 자 초
	4일차 (11.28)	금일도 보고서 작성(수정)을 계속했다. 보고서 상에서 데이터분석 시작과 끝을 보는 데에 이해할 수 있는 것이라는 판단하에 삭제하고, 쉽게 알 수 있는 것으로 데이터로 수정하였다. 제1기주지역부록 제3기주지역 등 나눠있는 둘째 용도를 사용하여 기주지역 면적별 CCTV설치 개소등 기초에 설치되어 있는 CCTV를 종려 세부적으로 분석하여 다양성을 늘렸다.	인 자 초
	5일차 (11.29)	금일 보고서를 확 라 작성하였다. 절차받는대로 혜안 시스템 표준모델에 올라갈 것이라고 하였다. 실시간으로 바뀌는 데이터와 새로 제공받는 데이터 때문에 계속해서 바뀌었지만, 오늘은 끝으로 최종 마무리 하였다. 이후 운영체제에서 상으로는 인구분석 혹은 대별로 시각화가 있는데 이를 과제로 반기전에 미리 학습하는 것이 좋을 것 같다. Java와 C 쪽을 학습해볼 계획이다.	인 자 초