

Avian Influenza

AIAI

Artificial Intelligence

고병원성 조류인플루엔자

확산 방지를 위한 지역별 전염 확률 예측 모델

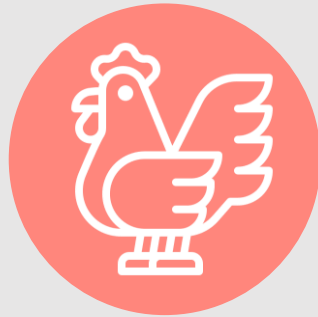
팀원 소개



이현지(팀장)
data analyst



이희주
data analyst



박지우
data analyst



하광림(멘토)



박세익(멘토)

목차

01

프로젝트 개요

- 1) HPAI 현황
- 2) 기획 배경
- 3) 기획 목표

02

개발 과정

- 1) 개발 프로세스
- 2) 세부 개발 내용

03

추후 일정

- 1) 추후 개발 일정

01

프로젝트 개요

- 1) HPAI 현황
- 2) 기획 배경
- 3) 기획 목표





고병원성 조류인플루엔자 (HPAI)

조류의 급성 전염병인 조류인플루엔자의 한 종류
제 1종 가축전염병으로 분류

1. 프로젝트 개요

1) HPAI 현황

과거 국내 AI 발생 현황 및 피해규모(출처 : 농림축산부)

구분	1차	2차	3차	4차
발생 시기	03.12.10~04.03.20 겨울(102일간)	06.11.22~07.03.06 겨울(104일간)	08.04.01~08.05.12 봄(42일간)	10.12.29~11.05.16 겨울~봄(139일간)
발생 범위	10개 시, 군	5개 시,군	19개 시,군	25개 시,군
피해 규모	392호, 5285천수 (1,531억원)	460호, 2800천수 (582억원)	1500호, 10204천수 (3,070억원)	286호, 6473천수 (822억원)

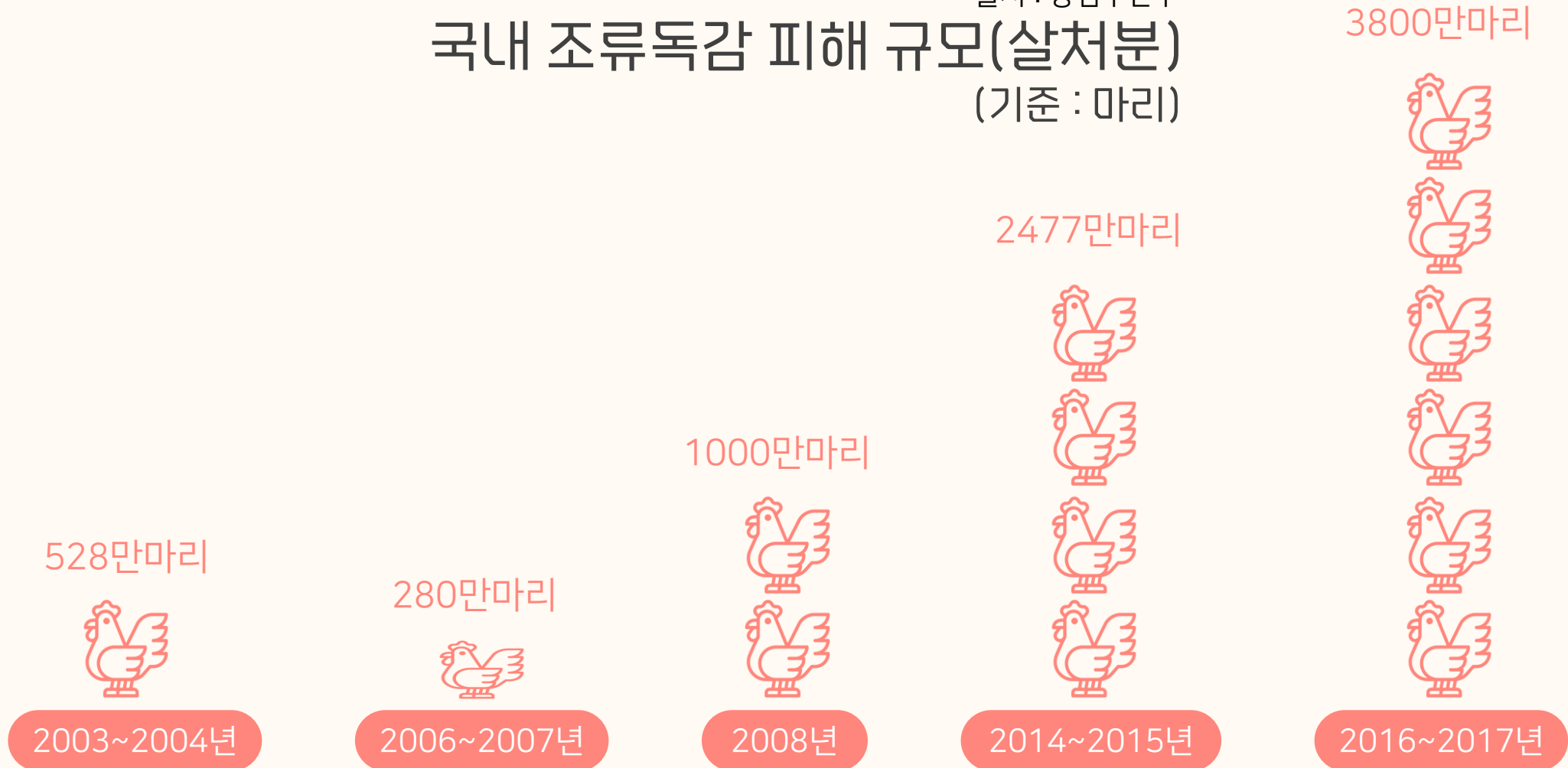
주기적, 지속적 발생

1. 프로젝트 개요

1) HPAI 현황

출처 : 농림수산부

국내 조류독감 피해 규모(살처분) (기준 : 마리)



1. 프로젝트 개요

2) 문제점

현재 AI 방역 대책 : 전국단위 일괄적 비상 방역

- 제주도에서 AI 발생시에도 전국 단위로 심각단계 경보 발생
- 전국단위 이동금지령
- 음성판정 농가 또한 3km 내 위치시 **무분별한 살처분 조치**



막대한 경제적 손실과 방역비용 발생



1. 프로젝트 개요

2) 문제점

AI 피해 규모 축소에 대한 필요성 요구

증가하는 AI 규모 및 피해 감소를 위한 효율적인 방역 체계 요구



과거 데이터를 통한 AI의 전염성 방향 분석 필요



1. 프로젝트 개요

3) 기획 목표

고병원성 조류인플루엔자
전염 예측을 통한 경제적 비용 절감

시군구 단위 지역별 전염 수준 예측 모델

지역별 전염 예측  선택적 살처분으로 인한 농가피해 최소화 목표

    등급별 위험도 시각화



02

개발 과정

- 1) 개발 프로세스
- 2) 세부 개발 내용



2. 개발 과정

1) 개발 프로세스

1. 데이터 수집



2. 분석 및 정제



3. 예측 모델



4. 웹 구현



2. 개발 과정

2) 세부 개발 내용 - WBS

9/16기준

모델별 성능 비교를 위해
모델링 진행중

웹 구현 및
독립변수 추가 예정

TASK NO.	TITLE & TASK	산출물	DUE DAY	상태
1	프로젝트 기획		@2021년 7월 12일 → 2021년 7월 30일	완료
1.1	계획			완료
1.2	정의 & 명세화			완료
2	기능 정의 & 명세화		@2021년 7월 23일 → 2021년 7월 30일	완료
2.1	계획			
2.2	기능 정의			
2.3	기능 명세화	기능 명세서		
3	개발			
3.1	데이터셋 검토 및 확보		@2021년 8월 2일 → 2021년 8월 13일	완료
3.2	데이터 분석		@2021년 8월 16일 → 2021년 8월 27일	완료
3.3	데이터셋 전처리		@2021년 8월 30일 → 2021년 9월 3일	완료
3.4	모델링		@2021년 9월 6일 → 2021년 9월 17일	진행중
3.5	웹 구현		@2021년 9월 20일 → 2021년 9월 30일	예정
4	테스트			
4.1	단위 테스트		@2021년 9월 17일 → 2021년 10월 5일	예정
4.2	통합 테스트		@2021년 9월 17일 → 2021년 10월 5일	예정
4.3	사용자 테스트		@2021년 9월 17일 → 2021년 10월 5일	예정
5	결과 보고			
5.1	산출물 정리		@2021년 9월 28일 → 2021년 10월 7일	예정
5.2	중간 결과 보고	PPT	@2021년 9월 15일	예정
5.3	최종 결과 보고	PPT	@2021년 10월 6일	예정

2. 개발 과정

2) 세부 개발 내용 - 발생 데이터

1. HPAI 발생 데이터 수집

번호	축주명	주소	축종	사육 두수	감염 두수	폐사 두수	검사 두수	진단명	진단일	발생일	기관	신규/ 추가	시스템 구분
1	신석	전라남도 장흥군 장 평면 양촌 리	오리-육용 오리	9500	0		100	고병원성조 류인플루엔 자	2021-04- 07	2021-04- 06	조류인플루 엔자연구진 단과	신규	수기등록
2	최은	전라남도 장흥군 장 평면 양촌 리	오리-육용 오리	15500	0		80	고병원성조 류인플루엔 자	2021-04- 07	2021-04- 06	조류인플루 엔자연구진 단과	신규	수기등록
3	홍지	전라남도 나주시 산 포면 등수 리	오리-육용 오리	19000	0		5	고병원성조 류인플루엔 자	2021-03- 24	2021-03- 23	조류인플루 엔자연구진 단과	신규	수기등록
4	박주	전라남도 장흥군 장 평면 기동 리	오리-육용 오리	11800	0		40	고병원성조 류인플루엔 자	2021-03- 23	2021-03- 22	조류인플루 엔자연구진 단과	신규	수기등록
5	이인	전라남도 장흥군 장 평면 기동 리	오리-육용 오리	20000	0		6	고병원성조 류인플루엔 자	2021-03- 22	2021-03- 21	조류인플루 엔자연구진 단과	신규	수기등록

농림수산검역본부의
2003년~2021년 발생 자료 사용

2. 시군구별 정제

주소	축종	사육수	진단일	발생일
강원도 원주시 귀래면 주포리	닭-산란계	170000	20210224	20210223
강원도 인제군 기린면 북리	닭-산란계	28500	20170107	20170103
강원도 철원군 갈말읍 강포리	닭-산란계	16000	20161212	20161209
강원도 철원군 갈말읍 강포리	닭-산란계	30000	20161203	20161130
강원도 춘천시 사북면 지촌리	닭-산란계	48	20080507	20080507
강원도 춘천시 사북면 지촌리	닭-산란계	48	20080507	20080507
강원도 횡성군 횡성읍 학곡리	닭-토종닭	13	20170109	20170106
강원도 횡성군 횡성읍 학곡리	거위-비분류	1300	20140617	20140613
경기도 고양시 덕양구 관산동	닭-토종닭	3000	20170307	20170303
경기도 고양시 덕양구 내유동	관상조류-비분류	900	20201229	20201228
경기도 과천시 문원동	야생조류-큰기러기	0	20140314	20140309
경기도 광주시 남한산성면 불당리	오리-기타	12	20160405	20160405
경기도 광주시 오포읍 양벌리	야생조류-왜가리	1	20150121	20150119
경기도 광주시 초월읍 신월리	닭-산란계	46000	20161222	20161218

불필요한 열 삭제, 날짜 정제
2003년~2020년 데이터 사용

2. 개발 과정

2) 세부 개발 내용 - 독립변수

1. 독립변수 선정

1. 기후 데이터 - 기상청
2. 철새 데이터 - 국립생물자원관
3. 거리 데이터 - 공공 데이터 포털, 카카오 API

2015~2020년 기준 데이터 사용
발생데이터 중 발생건수 종속변수 지정

2. 변수 추출

전국 시도, 시군구 기준 설정

1. 기후 데이터 : 평균 온도, 강수량
2. 철새 데이터 : 철새 수
3. 거리 데이터 : 최초발생지 거리

독립변수 추가 예정

2. 개발 과정

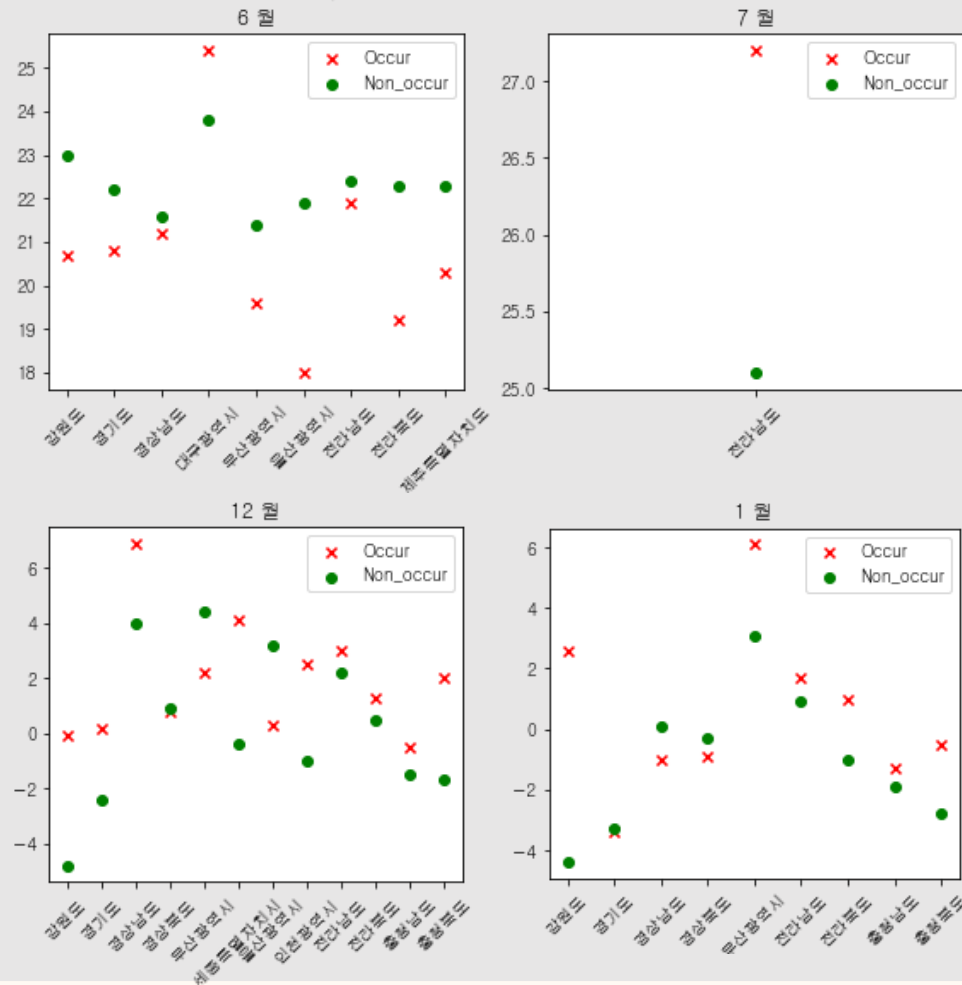
2) 세부 개발 내용 **기후 데이터 수집** - 기상청 일 자료 조회서비스 API 활용

온도

계절별 HPAI 발생일의 평균 온도와,
미발생일 평균 온도 차이가 있을 것으로 가정

Mann-Whitney U-test
검정 결과 유의수준 5% 이내로
변수 채택

여름철 낮은 온도,
겨울철 높은 온도 발생 확인



2. 개발 과정

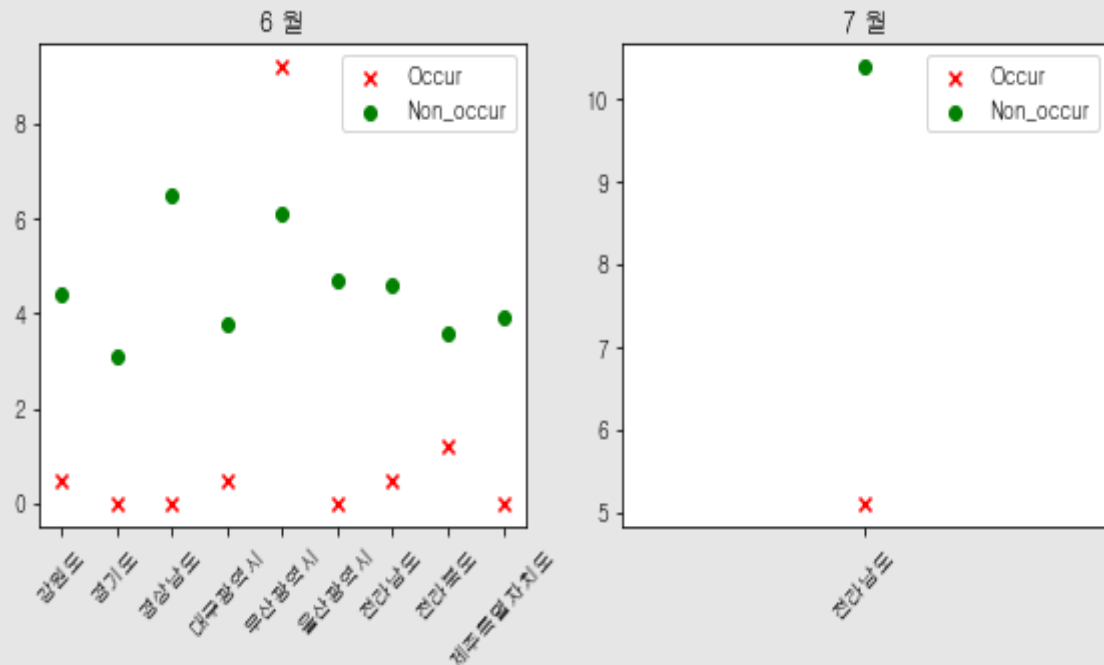
2) 세부 개발 내용 기후 데이터

강수량

여름 기준 HPAI 발생일의 강수량과 미발생일 강수량 차이가 있을 것으로 가정

Mann-Whitney U-test
검정 결과 유의수준 5% 이내로
변수 채택

여름철 적은 강수량일때
발생이 많은 것으로 확인



2. 개발 과정

2) 세부 개발 내용

기후 데이터 테이블

발생데이터
시계열 분석을 통한
사분위수 데이터로
신뢰구간 설정

25%~75%
신뢰 구간 리스크 부여

온도 강수량의 비율을
7:3으로 하여

위험도 레벨 차등 표현

시도	month	day	avg_temp(C)	daily_rain(mm)	temp_index	rain_index	risk
강원영동	1	1	1.2	0	0	0	0
강원영동	1	2	3.2	0	0	0	0
강원영동	1	3	3.4	0	0	0	0
강원영서	1	1	-4	0	0	10	5
강원영서	1	2	-2.2	0	0	10	5
강원영서	1	3	-2.7	0	0	10	5
경상남도	1	1	0.4	0	10	10	10
경상남도	1	2	1.8	0	10	10	10
경상남도	1	3	2	0	10	10	10
경상북도	1	1	-0.4	0	10	10	10
경상북도	1	2	1.2	0	10	10	10
경상북도	1	3	1.5	0	10	10	10
서울경기	1	1	-1.8	0	10	10	10
서울경기	1	2	0	0	10	10	10
서울경기	1	3	-0.5	0	10	10	10
전라남도	1	1	1.5	0.3	10	0	8
전라남도	1	2	3.3	0.1	10	0	8
전라남도	1	3	3.2	0	10	10	10
전라북도	1	1	0.4	0.4	10	0	8

2. 개발 과정

2) 세부 개발 내용 **거리 데이터 수집** - 공공 데이터 포털 전국 시군구 자료, 카카오 api 좌표 이용

거리

사전 검토 과정에서 거리가 필수 요소임을 확인

전국 발생지를 기준으로 전국 시군구 거리 계산

시도	시군구	1	2	3	4	5	6	7	8	...	97	98	99	100	101	102	103	104	105	106
강원도	인제군	44.7	50.6	57.1	78.8	79.3	85.5	93.8	95.4	...	269.1	273.6	275.1	279.4	279.9	293.6	297.2	320.9	358.1	389.3
강원도	횡성군	12.5	34.9	38.9	40.9	44.7	49.9	51.2	55.3	...	224.6	230.9	231.2	235.2	237.2	249.8	254.7	278.0	313.6	369.9
경기도	광주시	14.3	14.6	15.3	17.6	20.6	22.6	26.2	27.5	...	244.3	244.8	244.9	245.3	251.7	263.3	264.8	265.4	274.4	289.5
경기도	김포시	6.4	12.0	13.2	30.5	32.7	36.3	38.4	41.1	...	271.3	273.7	300.5	301.4	303.7	303.8	311.9	319.6	322.5	323.5
경기도	남양주시	14.6	14.8	16.3	17.5	17.6	21.4	21.5	25.0	...	254.7	256.5	256.8	257.1	261.4	275.2	275.7	277.1	283.9	286.3

2. 개발 과정

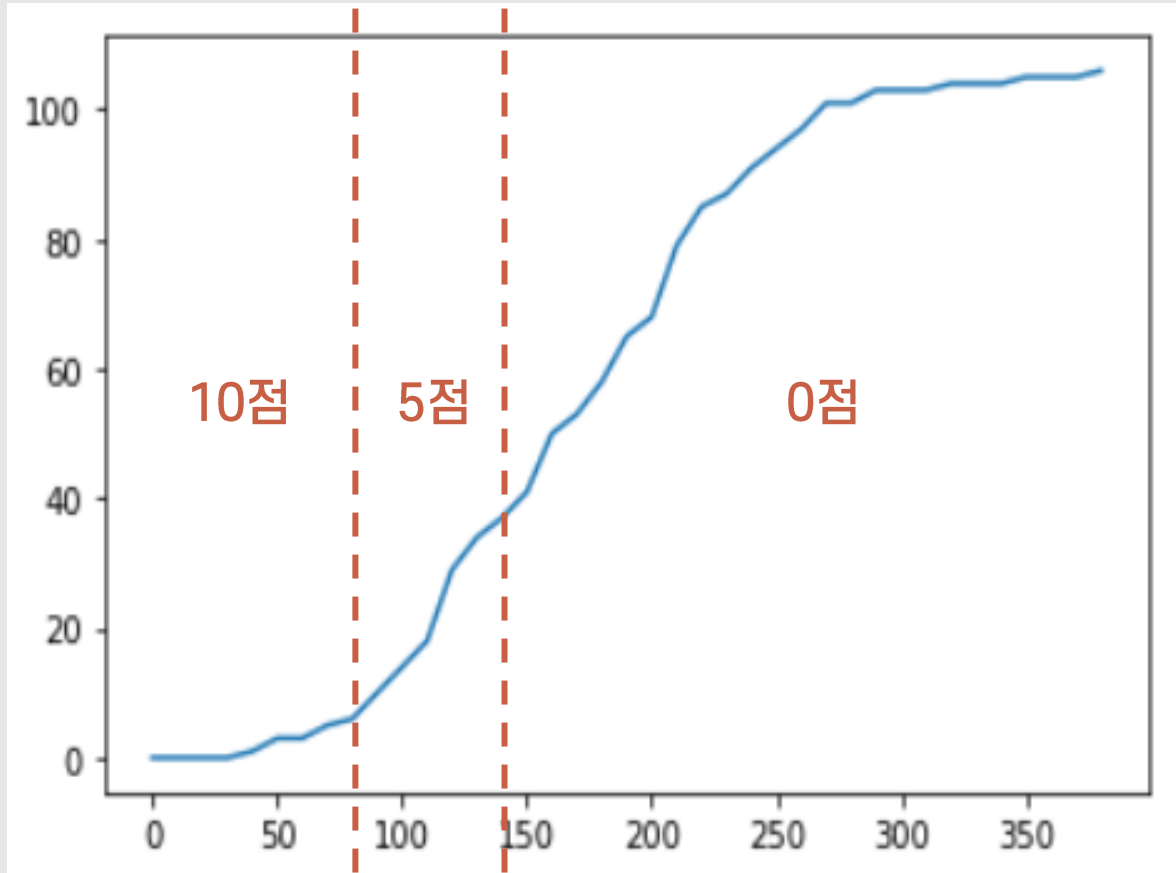
2) 세부 개발 내용 거리 데이터

Ex. 1월 최다 발생지(강원도 인제군) 기준

0~70km 10점 (25%)
80~140km 5점 (50%)
그 외 0점 부여

발생지 기준
월별 발생 건수 누적 분포도

같은 방식으로 2~12월 테이블 작성



2. 개발 과정

2) 세부 개발 내용

거리 데이터 테이블

월별, 거리별
리스크 부여

위험도 레벨 차등 표현

month	dist_range	cnt	cumulative	risk
1	80	3	48	5
1	90	5	53	5
1	100	5	58	5
2	80	5	33	10
2	90	12	45	5
2	100	7	52	5
3	80	5	33	10
3	90	12	45	5
3	100	7	52	5
4	80	8	45	5
4	90	5	50	5
4	100	4	54	5
5	80	8	31	5
5	90	1	32	5
5	100	4	36	5
6	80	3	5	10
6	90	2	7	5
6	100	3	10	5
7	80	2	24	10

2. 개발 과정

2) 세부 개발 내용 **철새 데이터 수집** - 국립생물자원관

철새수

철새 개체수가 많은 10~3월
HPAI 발생건수와
철새 개체수가 없는 4~9월
HPAI 발생건수 비교

10~3월의 발생건수가 4~9월의 2배 이상

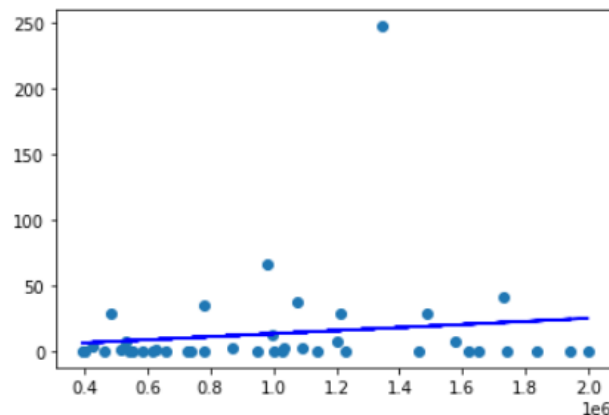
Mann-Whitney U-test
검정 결과 유의수준 5% 이내로
변수 채택

철새수와 발생건수의 선형관계 확인

```
In [27]: 1 regr=linear_model.LinearRegression()  
         2 regr.fit(x, y)
```

```
Out[27]: LinearRegression()
```

```
In [28]: 1 plt.scatter(x,y)  
         2 plt.plot(x,regr.predict(x), c='blue')  
         3 plt.show()  
         4  
         5 # x축=철새수  
         6 # y축=발생건수
```



2. 개발 과정

2) 세부 개발 내용

철새 데이터 테이블

시군구별 철새 수 기준
철새 숫자로 리스크 부여

예측발생건수에서
실제 발생건수 비율 확인
(발생건수 / 예측 발생건수)

위험도 레벨 차등 표현

month	시도	철새수	발생건수	예측 발생건수	발생건수/예측 발생건수	위험도
1	강원도	626130	2	9	0.22	0
1	경기도	1242626	55	16	3.44	3
1	경상남도	347682	11	6	1.83	1
1	경상북도	554884	2	8	0.25	0
1	광주광역시	49915	0	2	0	0
1	대구광역시	17006	0	2	0	0
1	부산광역시	411231	2	6	0.33	0
1	서울특별시	126690	0	3	0	0
1	세종특별자치시	55283	1	2	0.5	0
1	울산광역시	693134	0	10	0	0
1	인천광역시	214373	0	4	0	0
1	전라남도	1041056	37	14	2.64	2
1	전라북도	1773081	51	22	2.32	2
1	제주특별자치도	256564	0	5	0	0
1	충청남도	3370301	17	41	0.41	0
1	충청북도	212624	5	4	1.25	1
2	강원도	411800	1	6	0.17	0
2	경기도	962652	32	13	2.46	2
2	경상남도	134152	2	3	0.67	0
2	경상북도	325603	1	5	0.2	0
2	광주광역시	18287	0	2	0	0
2	대구광역시	2515	0	2	0	0

2. 개발 과정

2) 세부 개발 내용

최종 데이터 테이블

거리+기후+철새
데이터 테이블 합친 후

리스크 별 위험도 차등 부여

논문 등 관련 자료 검토 후
거리 : 철새 : 기후
5 : 3 : 2
비율 설정

독립변수 추가 예정

1	month	day	시군구index	철새수	최초발생기	평균온도	일강수량	위험도
2	1	1	0	126690	0	-1.8	0	7
3	1	1	0	126690	10	-1.8	0	7
4	1	1	0	126690	20	-1.8	0	7
5	1	1	0	126690	30	-1.8	0	7
6	1	1	0	126690	40	-1.8	0	7
7	1	1	0	126690	50	-1.8	0	7
8	1	1	0	126690	60	-1.8	0	7
9	1	1	0	126690	70	-1.8	0	7
10	1	1	0	126690	80	-1.8	0	4.5
11	1	1	0	126690	90	-1.8	0	4.5
12	1	1	0	126690	100	-1.8	0	4.5
13	1	1	0	126690	110	-1.8	0	4.5
14	1	1	0	126690	120	-1.8	0	4.5
15	1	1	0	126690	130	-1.8	0	4.5
16	1	1	0	126690	140	-1.8	0	4.5
17	1	1	0	126690	150	-1.8	0	2
18	1	1	0	126690	160	-1.8	0	2
19	1	1	0	126690	170	-1.8	0	2
20	1	1	0	126690	180	-1.8	0	2
21	1	1	0	126690	190	-1.8	0	2
22	1	1	0	126690	200	-1.8	0	2
23	1	1	0	126690	210	-1.8	0	2
24	1	1	0	126690	220	-1.8	0	2

2. 개발 과정

2) 세부 개발 내용 - 모델

정확도 & 재현율 향상을 위한 적용 검토 및 성능비교 단계

1. Random Forests

```
In [282]: train_x, test_x, train_y, test_y = train_test_split(train_data2, target_data2, test_size = 0.2, random_state = 42)
```

```
In [169]: clf = RandomForestClassifier(n_estimators=10, max_depth=10, random_state=42)
          clf.fit(train_x, train_y)
```

```
Out[169]: RandomForestClassifier(max_depth=10, n_estimators=10, random_state=42)
```

```
In [170]: predict1 = clf.predict(test_x)
          print(accuracy_score(test_y, predict1))
```

```
0.7570082387396403
```

2. 개발 과정

2) 세부 개발 내용 - 모델

정확도 & 재현율 향상을 위한 적용 검토 및 성능비교 단계

2. XG Boost

```
In [322]: train_X, test_X, train_Y, test_Y = train_test_split(train_data2, target_data, test_size = 0.2, random_state = 42)

In [323]: print(train_X.shape, test_X.shape)
(1958551, 8) (489638, 8)

In [324]: print(train_Y.shape, test_Y.shape)
(1958551, 1) (489638, 1)

In [325]: model = xgboost.XGBClassifier()
model.fit(train_X.values, train_Y.values)

In [376]: print("(pred == y_test) score: {0:.3f}".format((y_pred==test_Y['위험도'].values).mean()))
(pred == y_test) score: 0.913
```


2. 개발 과정

2) 세부 개발 내용 - 시각화

학습된 데이터로 시각화

Year, month,
day, 최초 발생지 입력



 자동 생성
(API 등을 이용한 실시간 데이터 호출)



Xgboost 모델을 통한
예측값인 위험도 생성

```
# 2021.09.01 시점 충청남도 공주시에서 HAPI 발생할 경우 테스트 data set (단일발생)
year=2021
month=9
day=1
place='충청남도 공주시'
```

year	month	day	시군구index	철새수	발생거리	평균온도	일강수량	위험도	code
2021	9	1	0	0	125.93	24.3	0	2.5	11110
2021	9	1	1	0	124.7	24.3	0	2.5	11140
2021	9	1	2	0	121.29	24.3	0	2.5	11170
2021	9	1	3	0	124.41	24.3	0	2.5	11200
2021	9	1	4	0	121.48	24.3	0	2.5	11215
2021	9	1	5	0	125.62	24.3	0	2.5	11230
2021	9	1	6	0	128.98	24.3	0	2.5	11260
2021	9	1	7	0	127.4	24.3	0	2.5	11290
2021	9	1	8	0	132.94	24.3	0	2.5	11305
2021	9	1	9	0	136.05	24.3	0	4.5	11320
2021	9	1	10	0	134.42	24.3	0	2.5	11350
2021	9	1	11	0	129.67	24.3	0	2.5	11380
2021	9	1	12	0	126.98	24.3	0	2.5	11410
2021	9	1	13	0	126	24.3	0	2.5	11440
2021	9	1	14	0	121.12	24.3	0	2.5	11470
2021	9	1	15	0	125.11	24.3	0	2.5	11500
2021	9	1	16	0	118.43	24.3	0	2.5	11530
2021	9	1	17	0	114.08	24.3	0	2.5	11545
2021	9	1	18	0	121.7	24.3	0	2.5	11560

2. 개발 과정

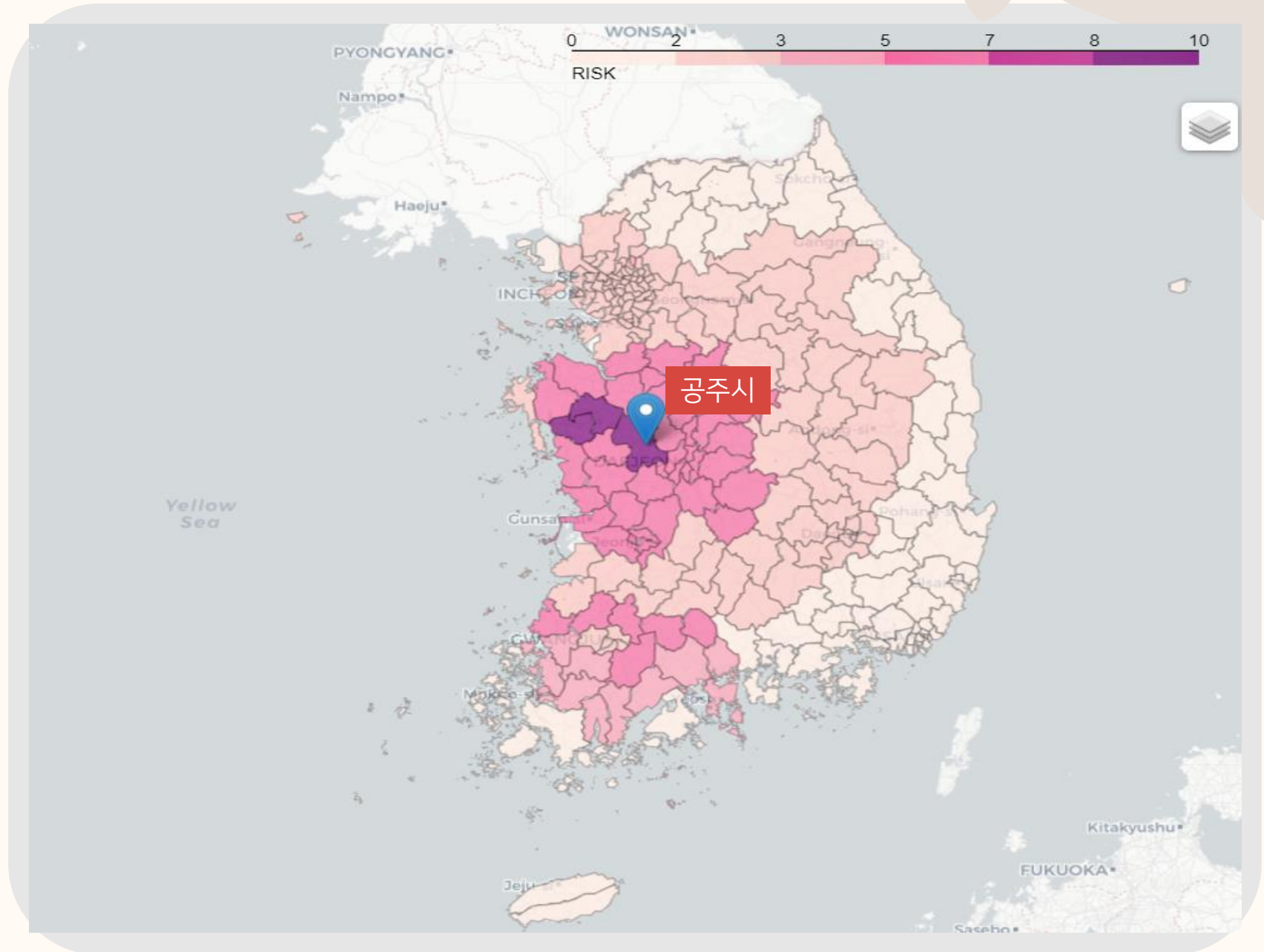
2) 세부 개발 내용 - 시각화

거리,기후,철새 변수로 모델링 후 시각화

2021/09/01
충청남도 공주시
발생 가정

위험도 시각화

추후 다중지역 발생
시각화 예정



2. 개발 과정

2) 세부 개발 내용 - 시각화

거리,기후,철새 변수로 모델링 후 시각화

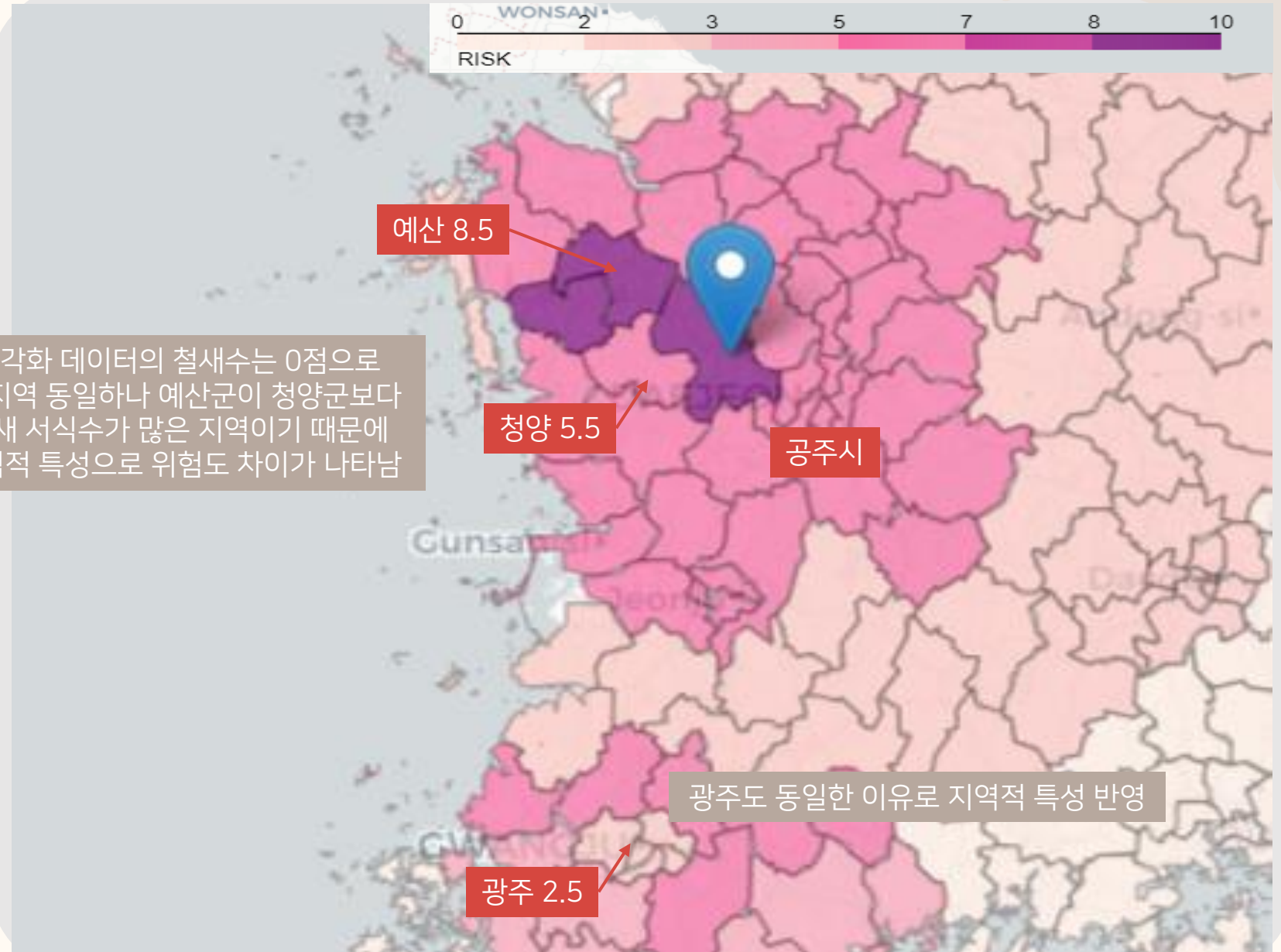
2021/09/01
충청남도 공주시
발생 가정

위험도 시각화

추후 다중지역 발생
시각화 예정

시각화 데이터의 철새수는 0점으로
두 지역 동일하나 예산군이 청양군보다
철새 서식수가 많은 지역이기 때문에
지역적 특성으로 위험도 차이가 나타남

광주도 동일한 이유로 지역적 특성 반영



03

추후 일정



03. 추후 일정

1) 추후 개발 일정

1. 독립변수 추가

독립변수 추가 예정

(농경지, 농장 수, 출하량,
통행량(KDB국가교통데이터베이스)
등 고려 중)

2. 모델링 & 시각화

추가된 독립변수 테이블
생성 및 추가 후
모델링

추가된 테이블로 시각화

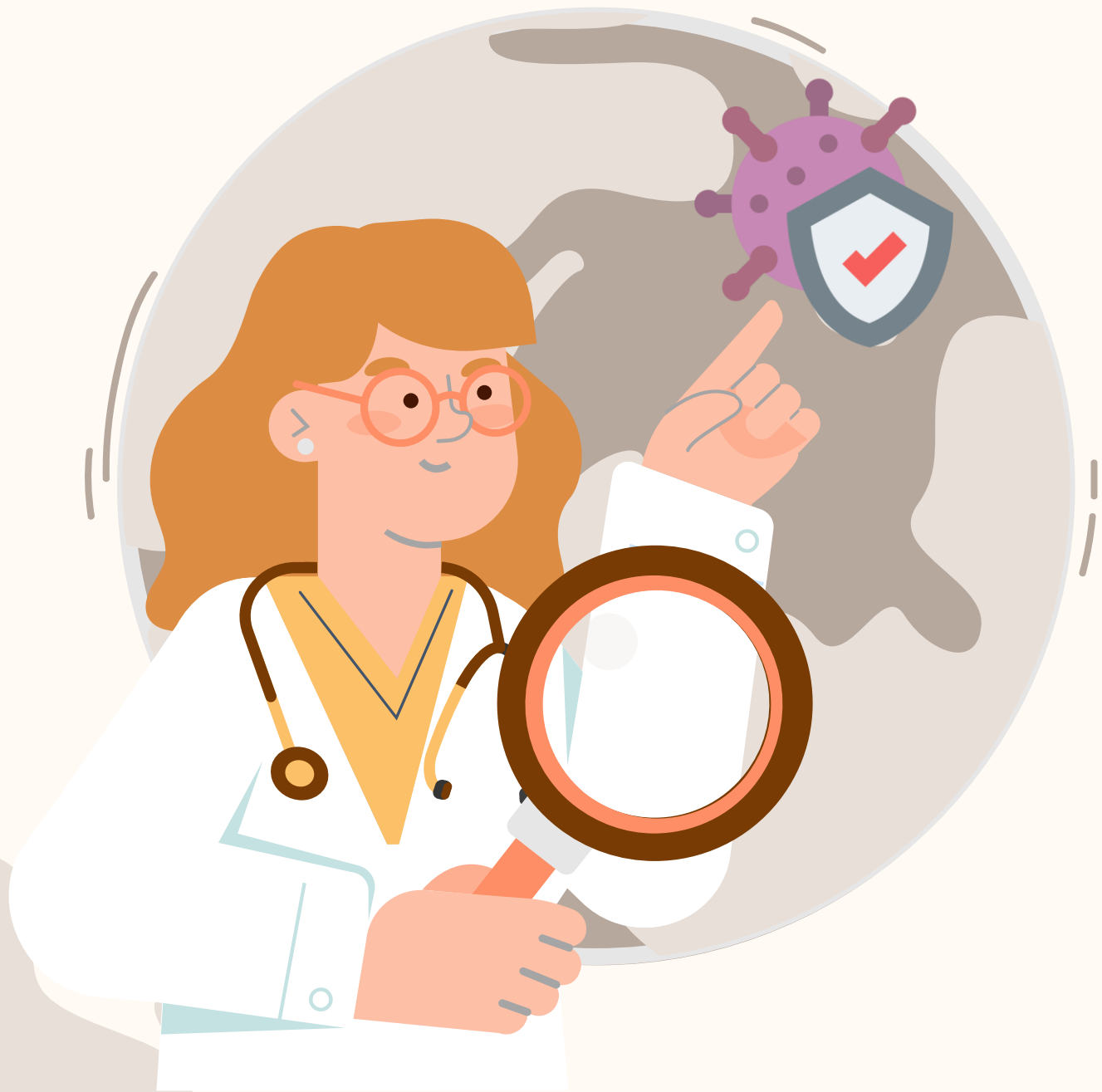
3. 웹 구현

최종 모델링 후
웹 구현 예정

Q & A

질문을주세요!





AIAI

감사합니다