**<요인 기본값 설정 이유>**

작물의 가격을 예측하기 위해 면적, 생산량, 물가지수, 물가상승률(총지수 전년동월비, 총지수 전년누계비, 신선식품지수 전년동월비, 신선식품 전년누계비), 날씨(평균기온, 평균최고기온, 평균최저기온, 평균상대습도, 월합강수량, 평균풍속, 합계일사량) 데이터를 모아 모델을 학습시켰다.

다만 모델학습을 하는 과정에서 피어슨 상관계수분석을 통해 상관관계가 낮은 요인들을 몇 개 발견하게 되었고, 이 요인을 배제하고 모델을 학습시켰을 때 오히려 정확도가 더 좋아지는 모습이 보여 몇 개의 요인(총지수 전년동월비, 신선식품지수 전년동월비, 신선식품지수 전년누계비, 합계일사량)을 제외하고 모델학습을 시키게 되었다.

**1. 면적**

면적은 전체적인 그래프를 보았을 때 최근 몇 년간 감소하는 추세를 보여주고 있다.

|  |
| --- |
|  |
| 배추상품 서울지역의 면적변화그래프 |

따라서 가장 최근인 2021년 데이터가 2022년에도 큰 변화가 없을 것으로 판단되어 2021년의 면적 값을 기본값으로 결정했다.

**2. 생산량**

작물의 생산량은 면적의 크기와 비례하는 모습을 보이며 그래프 또한 면적 그래프와 유사하게 최근 몇 년간 낮은 추세를 보여주고 있다.

|  |
| --- |
|  |
| 배추상품 서울지역의 생산량변화그래프 |

따라서 면적과 같이 가장 최근인 2021년 데이터가 2022년에도 큰 변화가 없을 것이라고 생각하여 2021년의 생산량을 기본값으로 결정했다.

**3. 물가지수**

물가지수는 2000년 1월부터 2022년 3월까지의 그래프를 봤을 때 값이 계속해서 크게 변하는 모습을 볼 수 있다.

|  |
| --- |
|  |
| 배추상품 서울지역의 물가지수변화그래프 |

어떤 값을 특정할 수 없기에 각 년도별로 그래프의 개형을 살펴보니 각 월마다 비슷한 값을 보여주는 걸 확인할 수 있었다. 6월을 기준으로 봤을 때 전체 년도(2000년~2021년)의 6월 평균과 최근 5개년의 평균, 2개년의 평균을 비교해보니 최근 5개년(2017년~2021년)의 6월 평균이 2021년 6월의 값과 가장 유사한 값을 보였다. 따라서 물가지수는 최근 5개년(2017년~2021년) 동월의 평균값으로 결정했다.

**4. 평균기온**

서울 : 1월을 기준으로 그래프의 최근 5년간(2017년~2021년)의 평균치가 -1.5정도이며, 2022월 1월 서울 평균기온이 -2.2이다. 전체 년도(2000년~2021년)의 평균을 냈을 때의 값이 -2.09로 전체 년도의 평균 값이 가장 비슷하기 때문에 평균기온의 기본 값은 전체 년도(2000년 ~ 2021년)의 평균 값으로 결정했다.

|  |
| --- |
|  |
| 서울지역의 평균기온을 그래프로 나타낸 결과 |

부산, 대구, 대전, 광주 : 그래프의 모양을 살펴봤을 때 최근 10년간의 그래프 모양이 유사하고, 평균 값이 2022년 1월과 가장 유사하기 때문에 최근 10년(2012년~2021년)의 동월 평균 값을 기본 값으로 결정했다.

**5. 평균최고기온**

서울 : 그래프의 모양이 온도가 상승하는 추세이다. 1월을 기준으로 2022년 1월의 값이 2.6이며, 전체 년도의 평균값(1.91)과 최근 3개년의 평균값(4.066), 최근 5개년(3.04), 최근 7개년(2.84)을 비교해봤을 때 최근 7개년의 값이 가장 2022년 1월의 값과 가장 유사하다. 따라서 평균최고기온의 기본 값은 최근 7개년의 평균 값으로 결정했다.

|  |
| --- |
|  |
| 서울지역의 평균최고기온을 그래프로 나타낸 결과 |

부산, 대구, 대전, 광주 : 그래프의 모양을 살펴봤을 때 최근 10년간의 그래프 모양이 유사하고, 평균 값이 2022년 1월과 가장 유사하기 때문에 최근 10년(2012년~2021년)의 동월 평균 값을 기본 값으로 결정했다.

**6. 평균최저기온**

서울 : 2022년 1월의 값이 -6.2이며, 최근 3개년의 평균값(-4.53)과 최근 5개년(-5.34), 최근 7개년(-5.47)을 비교해봤을 때 최근 7개년의 값이 가장 2022년 1월의 값과 가장 유사하다. 따라서 평균최저기온의 기본 값은 최근 7개년의 평균 값으로 결정했다.

|  |
| --- |
|  |
| 서울지역의 평균최저기온을 그래프로 나타낸 결과 |

부산, 대구, 대전, 광주 : 그래프의 모양을 살펴봤을 때 최근 10년간의 그래프 모양이 유사하고, 평균 값이 2022년 1월과 가장 유사하기 때문에 최근 10년(2012년~2021년)의 동월 평균 값을 기본 값으로 결정했다.

**7. 평균상대습도**

서울 : 그래프의 모양을 살펴보니 그래프의 모양이 살짝 내려가는 모양새이고, 10년을 기준으로 60을 넘는 데이터를 구분할 수 있게 되었기에 최근 10년간(2012년 ~ 2021년)의 평균 값을 기본 값으로 결정했다.

|  |
| --- |
|  |
| 서울지역의 평균상대습도를 그래프로 나타낸 결과 |

부산, 대구, 대전, 광주 : 몇 년간의 평균 값을 비교해본 결과 최근 5년간의 평균 값이 2022년 1월과 가장 유사하기 때문에 최근 5년(2017년~2021년)의 동월 평균 값을 기본 값으로 결정했다.

**8. 월합 강수량**

서울, 부산, 대구, 대전, 광주 : 그래프를 살펴보니 일정한 특징이 없는 것을 확인할 수 있었다. 하지만 2020년에는 비가 월등히 많이 내려 그래프에서 유난히 상승한 모습을 볼 수 있다. 서울을 기준으로 살펴보면 2020년을 제외하고 2015년~2021년 데이터의 평균 값을 살펴 봤을 때 값이 10.92이고, 전체 평균은 19.16이다. 최근 10개년의 평균값은 4.17이고, 2022년 1월이 5.5인 걸 보아 2020년의 값을 상쇄시키는 목적으로 범위를 넓게 잡아 최근 10년으로 하는 것이 옳다고 판단했다. 따라서 최근 10년(2012년 ~ 2021년)의 평균 값을 기본 값으로 결정했다.

|  |
| --- |
|  |
| 서울지역의 월합강수량을 그래프로 나타낸 결과 |

**9. 평균풍속**

서울 : 그래프의 모양이 거의 일정하고 반복되는 모습을 보여주고 있다. 데이터의 수치도 비슷한 모습을 보여주고 있기 때문에 전체 년도(2000년 ~ 2021년)의 평균을 기본 값으로 결정했다.

|  |
| --- |
|  |
| 서울지역의 평균풍속을 그래프로 나타낸 결과 |

부산, 대구, 대전, 광주 : 몇 년간의 평균 값을 비교해본 결과 최근 5년간의 평균 값이 2022년 1월과 가장 유사하기 때문에 최근 5년(2017년~2021년)의 동월 평균 값을 기본 값으로 결정했다.

**10. 총지수 전년 누계비**

총지수 전년 누계비는 절댓값이 주어지지 않은 비율 값이기 때문에 값이 불규칙하게 변하는 것을 확인할 수 있다.

|  |
| --- |
|  |
| 배추상품 서울지역의 총지수 전년 누계비 변화그래프 |

각 수치들을 뽑아보니 각 값들이 전월과는 큰 차이가 없다는 것을 알 수 있었다. 따라서 2022년 5월의 값을 기본값으로 결정하였다.