목차

1.팀 소개

2.프로젝트 소개

2-1.주제 선정 배경

2-2. 프로젝트 목표

3. 데이터 소개

4. 데이터 분석

4-1. 토양데이터 분석(맛데이터 분석 포함)

4-2. 기후 데이터 분석

-가설1. 연평균 온도가 높아지면, 당도가 높아질 것이다.

-가설2. 강수량이 많아지면, 당도/산도가 모두 낮아질 것이다.(일사량 얘기 포함)

-가설3. 야간 평균 온도가 높으면 당도/경도가 모두 낮아질 것이다.

-가설4. 일교차가 크면 당도/경도가 높을 것이다.

5. 결론

6. 참고문헌

안녕하세요, ‘사과사조’팀 사과의 맛에 기후 및 토양 요인이 미치는 영향 EDA

발표를 맡은 정별이라고 합니다.

**1.팀 소개**

저희 4팀은 모두 데이터 분석에 대한 열의가 가득한 사람들로 모였습니다. 각자 다음과 같은 역할 분담을 했습니다.

**2.프로젝트 소개**

**2-1.주제 선정 배경**

바로 프로젝트 소개로 넘어가겠습니다.

저희는 평범한 일상에서도 명절 때도 어김없이 빠지지 않는 국민 과일 ‘사과’에 대해 다루기로 했습니다. 그냥 평범한 사과가 아닌, 바로 ‘맛있는 사과’에 주목했는데요.

프로젝트의 가장 핵심 질문은 바로 ‘내가 농부라면, 맛있는 사과를 기르는데 어떤 요인을 고려해야 될까?’입니다. 다들 ‘어디 사과가 맛있다더라,,,’는 인식이 있으실 텐데, 저희는 그 중/ 서울 수도권의 사과 소비를 책임지는/ 신흥 강자 포천과, 전통적인 우리나라 사과의 대명사 대구의/ 군위를 설정했습니다. 중부 남부에서 가장 유명한 포천과 군위를 비교대상으로 분석을 진행하였습니다. 1m 12s

**2-2. 프로젝트 목표**

저희는 사과의 맛 요인에 대해, 기후와 토양 요인이 얼마나 어떠한 영향을 미치는지를 탐색적으로 분석(EDA)하고자 합니다. 여기서

v‘맛있는 사과’에 대해서는 아무래도 사과 종류에 따라 세부적인 기준이 다 달라서 수치적인 정의는 내리지 못했고, 맛에서 가장 많은 비중을 차지하는 ‘당도’에 초점을 맞춰서 다룰 계획입니다. 기본 통계분석과 상관계수 분석, 산점도 그래프 등의 시각화를 통해 진행하겠습니다.

**3. 데이터 소개**

(데이터 수집)활용할 데이터 소개입니다. 농촌진흥청에서 과실품질정보 데이터, 즉 사과의 당도, 경도, 산도와 토양데이터인 배수등급, Ph농도를 가져왔습니다. 기상자료 개방포털에서 연 강수량, 일교차, 기온, 일사량 등의 기상 데이터를 가져왔고 한국지질자원연구원의 ‘지오빅데이터 오픈플랫폼’에서 지열류량 데이터를 수집하였습니다.

(전처리) 전처리의 경우 이렇게 수집된 데이터를 연도와 지역에따라 사과 맛과 기후 데이터를 함께, 토양 데이터를 따로 분류하여 전처리했습니다. 분석으로 넘어가겠습니다.

**4. 데이터 분석**

**4-1. 맛데이터 분석**

먼저 맛데이터입니다. 당도, 산도, 경도로 이뤄져 있고, 두 지역을 비교해보았습니다. 단단함의 정도인 경도를 먼저 보시면, 한눈에 봐도 군위 값이 높습니다.

산도는 백분율, 당도는 브릭스 값으로 표현하는데, 두 지표에서는 0.01 차이도 중요한 의미를 갖습니다. 산도의 경우 평균과 중앙값 모두 군위가 0.23%로 포천보다 0.02% 높게 나타납니다. 여기서 산도가 높으면 신맛이 심하게 나니까 사과 맛이 떨어진다고 볼 수 있겠지만, 실제 0.5%를 넘어가지 않는다면 오히려 더 새콤달콤 맛있는 사과를 만들어줍니다.(책 <탐식생활>에의하면 산은 당을 잡는 동시에, 당은 산을 받쳐 준다. 상호 작용하는 두 맛이 조화를 이루어서 적당한 균형의 맛을 내는 것, 쉽게 말해 새콤달콤이 이상적인 사과의 맛이다.)

v가장 중요한 당도의 경우에는 2017~2019년에는 군위가 2020~2022년에는 포천이 상대적으로 크게 보이지만, (이전 슬라디ㅡ)실제 중앙값과 평균을 살펴보면 군위 14.5, 15.2에 포천 14.0, 13.2로 두 지표에서 모두 적지 않은 격차를 보여주네요.

정리하자면 사과의 맛은 당도, 산도, 경도로 구성되어 있고, 군위와 포천 각 7개년 지표들을 비교한 결과. 모든 지표에서 군위 사과가 더 맛있는 사과임을 알 수 있었습니다. 그렇다면 이렇게 맛있는 사과를 배출하는 군위만의 특징이 있을까요? 저희는 이를 분석하기 위해 우선 토양데이터를 살펴보았습니다. 5m 12 4m40s

**4-2. 토양데이터 분석**

~~토양데이터 입니다. 통계 분석과 공간 분석을 진행하였는데요. 크게 물 빠짐의 정도를 알려주는 배수등급, 땅의 산성도를 나타내는 ph농도, 땅 자체의 온도를 나타내는 지열류량 3가지를 분석하였습니다.~~

<먼저 배수등급입니다.> 3 매우양호에서 -3 매우불량까지 정수로 변환하여 그 분포를 분석하였는데요. 군위의 경우, 중앙값 2, 평균 2로 ‘약간 양호’한 모습을 보이지만, 포천은 중앙값 1, 평균 1로 그냥 ‘양호’한 모습을 보입니다. 표준편차는 유사하고요. 군위가 포천에 비해 배수 상태가 좋다는 뜻입니다.

<두번째로 땅의 산성도를 나타내는 ph농도입니다.> 군위의 경우 중앙값 6.55, 평균 6.51이고, 포천의 경우 중앙값 6.6, 평균 6.41로 그래프와 달리 두 지역간 전반적인 분포가 유사한 것을 알 수 있습니다. 적정 ph농도가 6.0~7.0인 것을 감안하면 군위와 포천의 사과 재배에 적합하다는 뜻이죠.

<세 번째는 땅의 온도를 나타내는 지열류량입니다.> 농가에서는 흔히 ‘지온’이라고 부르는데, 뿌리의 영양 흡수를 위해 중요한 요인입니다. 지온이 지나치게 높으면 양분 흡수량이 현격히 떨어집니다.

~~raw데이터가 지리데이터로 있어 그림으로 비교하였습니다.~~ 지도에서 가장 검은 곳은 약 22도로, 농작물의 적정온도인 20도 내외에 근접하고, 그 색깔이 하얘질 수록 적정온도로부터 증가, 멀어지는 것입니다. 보시다시피 표시된 군위에 비해 포천의 지열류량이 높은 것을 알 수 있습니다.

이상 포천과 군위의 맛데이터를 분석하고 이에 영향을 미치는 토양 데이터를 살펴보았습니다.

(경도, 산도, 당도 모두 군위의 지표가 포천에 비해 높아 더 맛있는 사과임을 알 수 있었고, 그 이유를 토양 요인에서 확인하고자 했는데요. 그 결과 배수등급, ph농도, 지열류량에서 모두 군위가 포천에 비해 사과 재배에 적합한 환경임을 알 수 있었습니다.)

**4-3. 기후 데이터 분석**

그렇다면 기후 데이터에서도 과연 이러한 양상을 보여줄까요? 기후 데이터의 경우, 해당 변수들이 많아 가설을 중심으로 분석을 진행하였습니다. 특히 이전에 진행했던 단순 수치나 통계 분석에서 나아가, 독립변수와 종속변수 간의 직접적인 관련성을 보여주는 ‘상관계수’를 중심으로 가설을 검증하였습니다.

**-가설1. 강수량이 많아지면, 당도가 낮아질 것이다.(일사량 얘기 포함)**

v강수량/강수일수입니다. 지도에서 포천이 속한 하늘색 부분이, 군위가 속한 노란색에 비해 강수량과 강수일수가 많은데요. 실제 통계에서도 포천의 강수일수가 군위에 비해 3일이나 더 많고, 강수 때문인지는 몰라도 포천의 사과 당도가 군위에 비해서 낮은 모습을 보여줍니다. 그래서 저희는 강수량이 많아지면, 당도가 낮아질 것이라는 가설을 설정했습니다.

v실제로 강수량과 당도 사이에 상관계수를 확인한 결과 포천에서는 가설과달리 0.86이라는 매우 높은 양의 상관관계가 나타나는데요. 이에 저희는 그 이유를 찾기 위해 추가적으로 조사를 했고, 1년의 강수량/강수 일수에 관계없이 수확 2주일 전부터 수확날까지 일사량이 좋으면 달고 맛있는 과일이 되는 것을 알게 되었습니다. 이에 인사이트를 얻어, 두 지역의 수확 시기를 고려해 데이터 기간을 좁혀 분석을 진행하였는데요.

v기간을 좁히니 강수량과 당도의 관계는 두 지역에서 모두 의미를 찾지 못했고, 오히려 일사량과 당도 사이의 양의 상관관계를 알 수 있었습니다.

v즉, 군위와 포천 모두, 1년이 아닌 수확 직전 시기의 일사량이 해당 시기의 강수량에 비해 더 큰 영향을 미친다는 인사이트를 도출하게 되었습니다.

**-가설2. 연평균 온도가 높아지면, 당도가 높아질 것이다.**

v일반적으로 따뜻한 지역의 과일이 더 맛있고 달다고 알려져있는데요. 실제 위도 차이에서 알 수 있듯이 군위가 포천보다 평균적으로 1도 정도 높습니다. 그 영향인지 당도 또한 군위 사과가 더 높기에, ‘연평균 온도’와 ‘당도’의 양의 상관관계를 가설로 설정하였습니다.

v하지만 히트맵을 통해 상관계수를 분석한 결과 둘 다 절댓값이 0.2~0.3으로 낮은 수준을 보여주었습니다. 이는 군위와 포천, 각각의 당도와 평균기온 그래프를 통해 확인할 수 있는데요. 군위에서는 2017, 2018년 평균기온이 적정온도 8~11도에 근접한 탓에 당도가 높게 나온 것으로 보이지만,

v군위와 포천 모두 그 추이가 일정하지 않음을 알 수 있습니다. 그래서 가설의 결론은 ‘알 수 없다’입니다.

**-가설3. 야간 평균 온도가 낮으면 당도가 모두 높아질 것이다.**

v농촌 진흥청에 따르면 7~9월 야간 평균 온도는 당의 축척에 있어 중요한 요인입니다. 그 적정 수준은 8도로 이에 근접할 수록 당도가 높게 나오는 것이죠. 실제 군위와 포천의 야간 평균 기온 추이를 보면 비슷한 양상을 보이지만, 수치 자체는 전반적으로 둘다 8도를 넘고 있습니다. 이에 저희는 ‘야간평균 기온이 낮을 수록 적정기온에 가까워지니 당도가 높아질 것이다’는 가설을 설정하였습니다.

v실제 히트맵을 통해 상관관계를 분석한 결과 두 지역에서 다른 모습을 보여주었는데요. 먼저 군위입니다. 야간평균 온도와 당도의 관계가 -0.85라는 매우 높은 상관계수로 나타나네요. 실제 그래프 상에서도 회색선 야간평균 기온이 상승할 때, 빨간선 당도가 하락하는 추세를 명확히 보여주죠.

v다음은 포천인데요. 야간평균 온도와 당도의 관계가 +0.52의 상관관계를 보여줍니다. 가설과 반대되는 결과이죠. 이는 가설 설정 단계에서 뭉뚱그려 넘어간 포천의 실제 야간평균 온도에서 그 답을 찾을 수 있습니다.

v야간의 적정온도는 8도라고 말씀드렸는데, 포천의 온도 범위는 처음에 언급한 8이상이 아니었고 실제 7.7에서 8.5 정도였습니다. 이는 8.3에서 9.5의 범위를 갖는 군위와 큰 차이가 있는 것입니다. 즉, 포천의 야간 평균 온도는 전반적으로 8보다 낮거나 가깝기에, 그 수치가 증가할 수록 적정 온도에 가까워집니다. 그렇기에 온도가 높아질 수록 당도가 높아지는 양의 상관관계를 보이는 것이죠.

**-가설4. 일교차가 크면 당도가 높을 것이다.**

v야간 평균온도와 깊은 관련성이 있는, 일교차의 경우 사과 재배에 중요한 요인입니다. 12~17도가 재배에 적합한 수준으로 통계상으로는 두 지역 모두 적합해 보입니다.

v먼저 군위의 경우 일교차와 당도의 관계가 0.62의 양의 상관관계를 보여줍니다. 실제 그래프에서도 두 변수가 동일한 방향성을 나타내죠. 이는 높은 수준의 야간평균기온과 직접적 관련성이 있습니다. 바로 이전 가설에서, 군위의 야간평균기온이 낮을 수록 8도인 적정온도에 부합하여 당도가 높아진다고 말씀드렸습니다. 근데 이 야간평균기온이 낮아질 수록, 최저기온이 낮아질 수록 일교차가 커집니다.

정리하자면 군위에서 야간평균기온이 낮아지면 당도는 올라가는데, 동시에 야간평균기온이 낮아지면 일교차도 올라갑니다. 자연스럽게 당도와 일교차 서로 올라갈 수 밖에 없는 것이죠. 이와 달리 포천에서는 유의미성을 찾기 어려웠구요.

**5. 결론**

지금까지 포천과 군위의 맛있는 사과에 영향을 미치는 요인을 분석하고자, 토양과 기후의 여러 요인을 살펴보았습니다. 토양데이터의 경우 수치와 통계 분석을, 기후데이터의 경우 상관분석을 중심으로 진행하였구요.

사과의 맛 자체는 군위가 포천보다 나음을 알 수 있었고, 실제 군위의 토양 요인이 포천에 비해 양호함을 확인하였습니다. 이에 비해 기후 요인의 경우, 군위는 주로 기온 환경이 포천은 주로 일사량이 사과 당도에 큰 영향을 미쳤습니다.

이를 통해 사과 재배에 적합한 장소 선택을 위해서는 일부 공통된 특징은 존재하지만, 지역에 따라 적합한 환경변수를 고려해야 한다는 인사이트를 얻었습니다.

‘탐색적’ 데이터분석인 만큼 내용 전개가 매끄럽지 못했지만, 사과 재배에 어떤 요인들이 중요하고 다양한 변수를 고려해야 한다는 인사이트 도출에 의의가 있었다고 생각합니다.

긴 발표 들어주셔서 감사합니다. 이상입니다. 17m 39s

6. 참고문헌