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 ‘맛있는’ 사과와 이에 영향을 미치는 요인 EDA

발표를 맡은 정별이라고 합니다.

1.팀 소개

저희 4팀은 관심 분야는 조금씩 달라도, 모두 데이터 분석에 대한 열의가 가득한 사람들로 모였습니다. 각자 다음과 같은 역할 분담을 했어요.

프로젝트 소개입니다.

2.프로젝트 소개

2-1.주제 선정 배경

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

프로젝트의 가장 핵심 질문은 바로 ‘내가 농부라면, 맛있는 사과를 재배하기 위해서 어떤 요인을 고려해야 될까?’입니다. 다들 어디 사과가 맛있다더라,,,는 인식들이 있으실 텐데, 중부 지방과 남부 지방에서 가장 유명한 포천과 군위를 비교대상으로 하여 분석을 진행하였습니다.

2-2. 프로젝트 목표

저희는 사과의 맛 요인에 대해, 기후와 토양 요인이 얼마나 어떠한 영향을 미치는지를 탐색적으로 분석(EDA)하고자 합니다. 기본 통계분석과 상관계수 분석, 산점도 그래프 등의 시각화를 통해 진행하겠습니다.

3. 데이터 소개

활용할 데이터 소개입니다. 농촌진흥청에서 과실품질정보 데이터, 즉 사과의 당도, 경도, 산도를 가져왔고, 기상자료 개방포털에서 연 강수량, 습도, 일교차, 기온, 일사량 등의 기상 데이터를 가져왔습니다. 농촌진흥청의 ‘농사로’ 홈페이지에서 배수등급, ph농도를, 한국지질자원연구원의 ‘지오빅데이터 오픈플랫폼’에서 지열류량 데이터를 수집하였습니다.

4. 데이터 분석

4-0. 맛데이터 분석

먼저 맛데이터입니다. 당도, 산도, 경도로 이뤄져 있고, 연도별로 두 지역을 비교해보았습니다. (군위vs포천 사과품질\_경도)단단함의 정도인 경도를 먼저 보시면, 한눈에 봐도 군위 값이 높은 걸 알 수 있죠.

산도는 백분율, 당도는 전용 지표인 브릭스 값으로 표현하는데, 경도와 달리 두 지표에서는 0.01 차이도 중요한 의미를 갖는다고 생각하시면 됩니다. 소수점 단위에서 사과의 맛이 변화가 크게 나타나는 것이죠. 산도의 경우 평균과 중앙값 모두 포천이 0.23%로 군위보다 0.02% 높게 나타납니다. 여기서 산도가 높으면 신맛이 심하게 나니까 사과 맛이 떨어진다고 볼 수 있겠지만, 실제 0.5를 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.)

가장 중요한 당도의 경우에는 2017~2019년에는 군위가 2020~2022년에는 포천이 상대적으로 크게 보이지만, 실제 중앙값과 평균을 살펴보면 군위 14.5, 15.2에 포천 14.0, 13.2로 두 지표에서 모두 적지않은 격차를 보여주네요.

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

4-1. 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 ph농도 요인은 사과 재배에 적합하다는 뜻이죠.

<세 번째는 땅의 온도를 나타내는 지열류량입니다.> raw데이터가 지리데이터로 저장되어 있어 정확한 수치보다는 그림으로 비교하였습니다. 지도에서 색깔이 하얄수록 지열이 높고 검을수록 지열이 낮은데, 표시된 군위에 비해 포천의 지열류량이 높은 것을 알 수 있습니다.

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

뚜렷한 경향성 안보임

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