Hallo Anne, Feedback op deze versie van je challenge:

1) In een notebook /altijd/ beginnen met context: Waar hoort dit bij? (AI4 semester Fontys ICT, individuele challenge, je eigen naam, maart/april 2024...) Maar ook: wat wil je bereiken? Waar komt je data vandaan? Wat zit er in die data? (scope, tijdsbereik, etc.) Wat is de verwachte bruikbaarheid/kwaliteit van die data voor je project? Is er evt. al voorwerk gedaan aan de data in een ander notebook? Een indicatie welke data-preparatie en/of ML methodes je verwacht te gaan gebruiken is ook een goed idee.

2) Bij jupyter notebooks is het bad practice (Een slechte gewonte) om comments in je code te zetten (behalve eventueel heel technische opmerkingen). Good practice (goede gewoonte): losknippen in kleine stukjes code, met text-cellen tussendoor ipv comment lines.

3) Dit notebook bevat te weinig text-cells. De kopjes zijn wel suggestief, maar: Vuistregel: zonder de code te zien, alleen door de text-cells te lezen, moet ik als lezer je hele gedachtengang kunnen volgen.

4) Bij 'all train data' ga je veel te snel over op weggooien van null values. Eerst goede EDA doen. (O.a. voor alle 'object' features: distinct values bekijken, etc. Is er reden voor one-hot encoding? Over alle features moet je even een mening vormen (en opschrijven in text-cells!!)

5) Bij decision tree beperk je je op voorhand tot een subset van de features. Op grond waarvan? Lage correlatie kan misleidend zijn. Waarom voorspel je daar trouwens de temperatuur? Welke relevantie heeft dat voor je project?

6) Warnings en errors altijd eerst oplossen voor je een notebook oplevert.

7) Waarom niet even de distinct values laten zien van 'statistical cause'? (En van andere categorische features.)

8) Je vertaalt categorische features naar getallen. Is dat zinvol? Is er verband tussen de ordening die zo ontstaat en de betekenis van de oorspronkelijke waardes?

9) Telkens als je een result hebt, weer even een text-cell met conclusies. Zie je wat je verwachtte te zien? Zie je iets wat je niet verwachtte, en waar je dieper op in gaat?