

# Normalformer

Av Katrine :D



Brnavn	Navn	Etternavn	Adresse	Kurskode	Tittel	Beskrivelse	AntSP	Kara
evgenit	Evgenij	Thorstensen	Addr1	IN2090	Databaser	EnBeskr...	10	B
peternl	Petter	Nilsen	Addr2	IN2090	Databaser	EnBeskr...	10	A
evgenit	Evgenij	Thorstensen	Addr1	IN2080	Beregn...	Descr...	10	A
leifhka	Leif H.	Karlsen	Addr3	IN2090	Databaser	EnBeskr...	10	B
leifhka	Leif H.	Karlsen	Addr3	IN3110	Program...	EnBeskr2...	5	C



# Hvorfor er dette en dårlig tabel?





# Normalformer er viktig for å unngå duplikater

- Lagrer samme data mange ganger, som tar opp mye plass
- Ineffektive spørringer
- Ved oppdateringer må informasjon som allerede ligger inne skrevet på nytt
- Mange NULLer når det er kolonner som ikke er relevante for andre kolonner
- Alle duplikater må bli oppdaterte ved endringer
- Ved sletting slettes også urelaterte data



# En god tabell

- Har data om en entitet i sin egen tabell
- Har relasjoner mellom entiteter som referanser
- For dette har vi normalformer, kriterier som skal oppfylle kravene for gode tabeller





Karakter		
Brnavn	Kurskode	Kara
evgenit	IN2090	B
peternl	IN2090	A
evgenit	IN2080	B
leifhka	IN2090	B
leifhka	IN3110	C

## Funksjonell avhengighet

- Et attributt A er funksjonelt avhengig av en mengde attributter X hvis det bare kan finnes en verdi av A for hver mengde verdier av attributtene i X
- Skrives  $X \rightarrow A$ . Kalles FD (functional dependency)
- Brnavn, Kurskode  $\rightarrow$  Karakter
- FDer forteller oss om hvilke data som hører sammen, og hva de hører til



## Hvordan de henger sammen

- $\{A, B\} \rightarrow \{C, D\}$  kan skrives:
- $A, B \rightarrow C, D$
- Hvis det er bokstaver kan vi også skrive  $AB \rightarrow CD$
- Når det kommer til flere avhengige, er:
- $AB \rightarrow CD \iff AB \rightarrow C \text{ og } AB \rightarrow D$
- $A \rightarrow B$  og  $B \rightarrow C$  betyr at  $A \rightarrow C$  (transivitet)
- Trivielle FDer som  $AB \rightarrow B$  ignoreres, siden de ikke gir oss noe nytt





## Kjente begreper og logikk :)

- En supernøkkel er enhver mengde attributter som er unike for relasjonen. Dette kan dermed inkludere hvert eneste attributt i relasjonen
- En kandidatnøkkel er en minimal supernøkkel. Som at {personnr, navn} er en supernøkkel, mens {personnr} er en kandidatnøkkel
- FDene sier hvilke supernøkler og kandidatnøkler vi har!





## $R(A, B, C)$

- Dersom  $R$  er en relasjon med attributter  $X$ , så vil:
- $Y$  mengde i  $X$  være en supernøkkel for  $R$  hvis  $Y \rightarrow X \setminus Y \implies Y \rightarrow X$
- $R(A) \rightarrow R(A, B, C) \setminus R(A) \implies R(A) \rightarrow R(B, C)$
- $Y$  mengde i  $X$  er en kandidatnøkkel hvis  $Y$  er en minimal supernøkkel
- For å sjekke om  $X$  er en supernøkkel, sjekk om alt er avhengig av  $X$ .  $R(A)$  er en supernøkkel fordi  $R(A, B)$  er avhengige av den



- ◆ **Tillukningen**  $X^+$  av  $X$  på en mengde FDer er mengden attributter som er funksjonelt avhengige av  $X$
- ◆ Hvis  $X \rightarrow A$ , så er  $A \in X^+$  sant
- ◆ Hvis  $A \notin X^+$ , så er ikke  $X \rightarrow A$  sant
- ◆ Tillukningen kan regnes ut ved å bruke FDene om og om igjen:
  - ◆ sett  $X^+ = X$
  - ◆ så lenge  $X^+$  forandres:
    - ◆ finn en FD  $Y \rightarrow Z$  med  $Y \subseteq X^+$
    - ◆ sett  $X^+ = X^+ \cup Z$

## Tillukning av $R(A, B, C, D, E)$

- $A \rightarrow B, AC \rightarrow E, D \rightarrow A$
- $A^+ = AB, AC^+ = ABCE$ , hva med  $D$ ?
- $D^+ = DA, DA^+ = DAB, DAC^+ = DABCE$
- Med nøkkelen  $ACD$  kan vi komme til alle attributter i relasjonen
- Regler for kandidatnøkler:
  - Hvis  $X$  bare forekommer på venstresiden, så må den være en del av kandidatnøkkelen.  $D$  og  $C$ , men ikke  $A$
  - Hvis  $Z$  bare forekommer på høyresiden kan den ikke være en del av kandidatnøkkelen.  $B$  og  $E$
- Fordi vi kan nå  $A$  med  $D$  vil  $ACD$  være en supernøkkel, men  $CD$  vil være en kandidatnøkkel
- Sjekk flere kombinasjoner som lar deg nå alle attributter, og sammenlign hvem som lar deg nå alle med færrest attributter





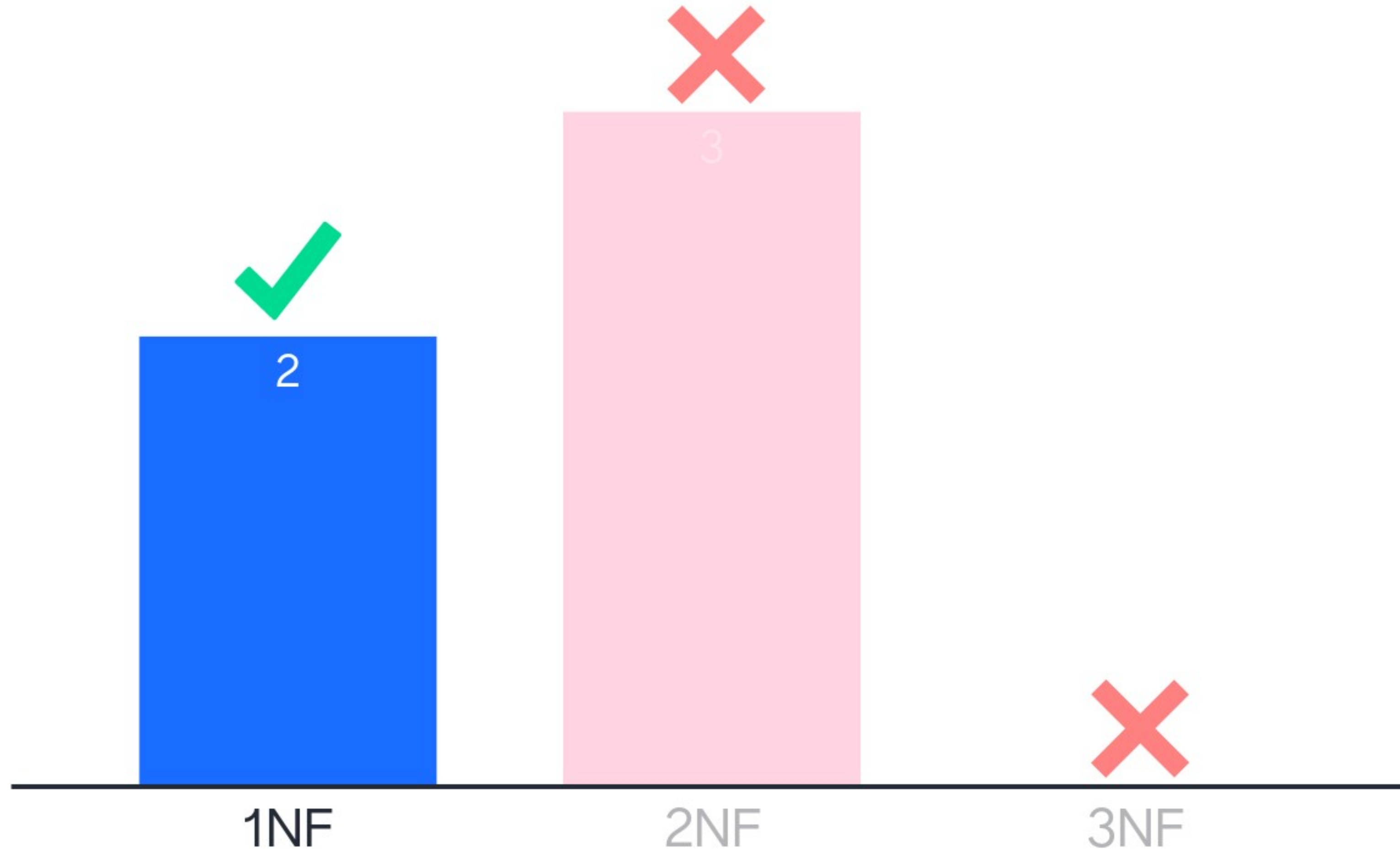
## BCNF er mengde i 3NF er mengde i 2NF er mengde i 1NF

- BCNF er best :) Her kommer krav
- 1NF: Atomære og inneholder bare en verdi. Vi antar at den alltid er oppfylt
- 2NF: Oppfyller 1NF og at alle attributter som ikke er nøkkelattributter, ikke er FD av en delmengde av en kandidatnøkkel
- NOTE: Nøkkelattributt er et attributt som er en del av en kandidatnøkkel, så om  $R(ABC)$  er en kandidatnøkkel, så kan ikke  $AB \rightarrow D$  og  $BC \rightarrow D$
- 3NF: Oppfyller 2NF og alle ikke-nøkkelattributter kun er avhengige av kandidatnøkler, hvor  $R(X)$  og  $X \rightarrow Y$  og ikke  $X \rightarrow Z$  og  $Z \rightarrow Y$
- BCNF: Oppfyller 3NF og alle attributter kun er avhengig av en kandidatnøkkel
- I 3NF kan du ha kandidatnøkler  $R(AB)$  og  $R(BC)$ , men for å oppfylle BCNF må alle attributter bare være direkte avhengig av  $R(AB)$
- Høy NF = få anomaliteter, men også mange relasjoner. Northwind og filmdatabasen er begge på BCNF

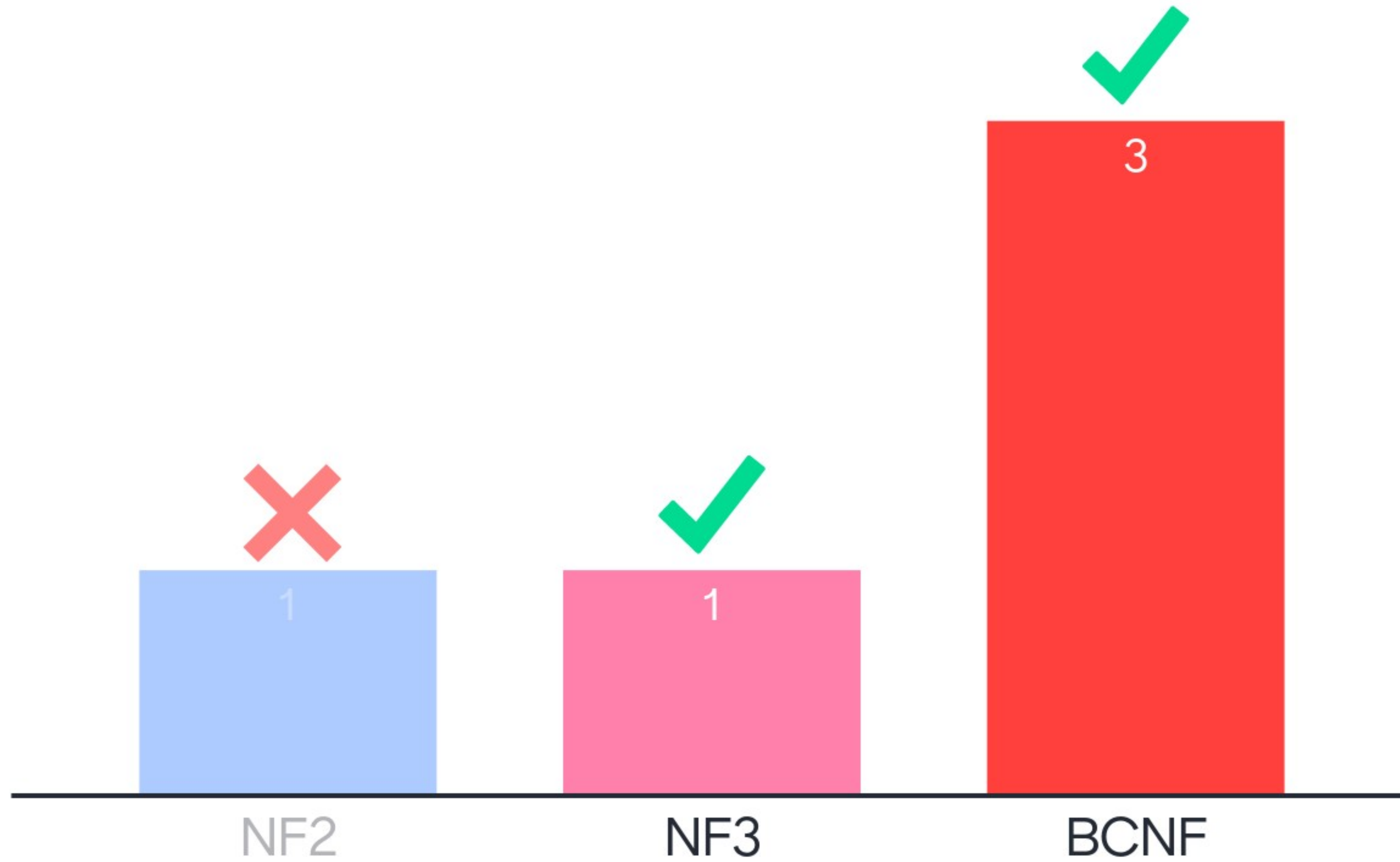




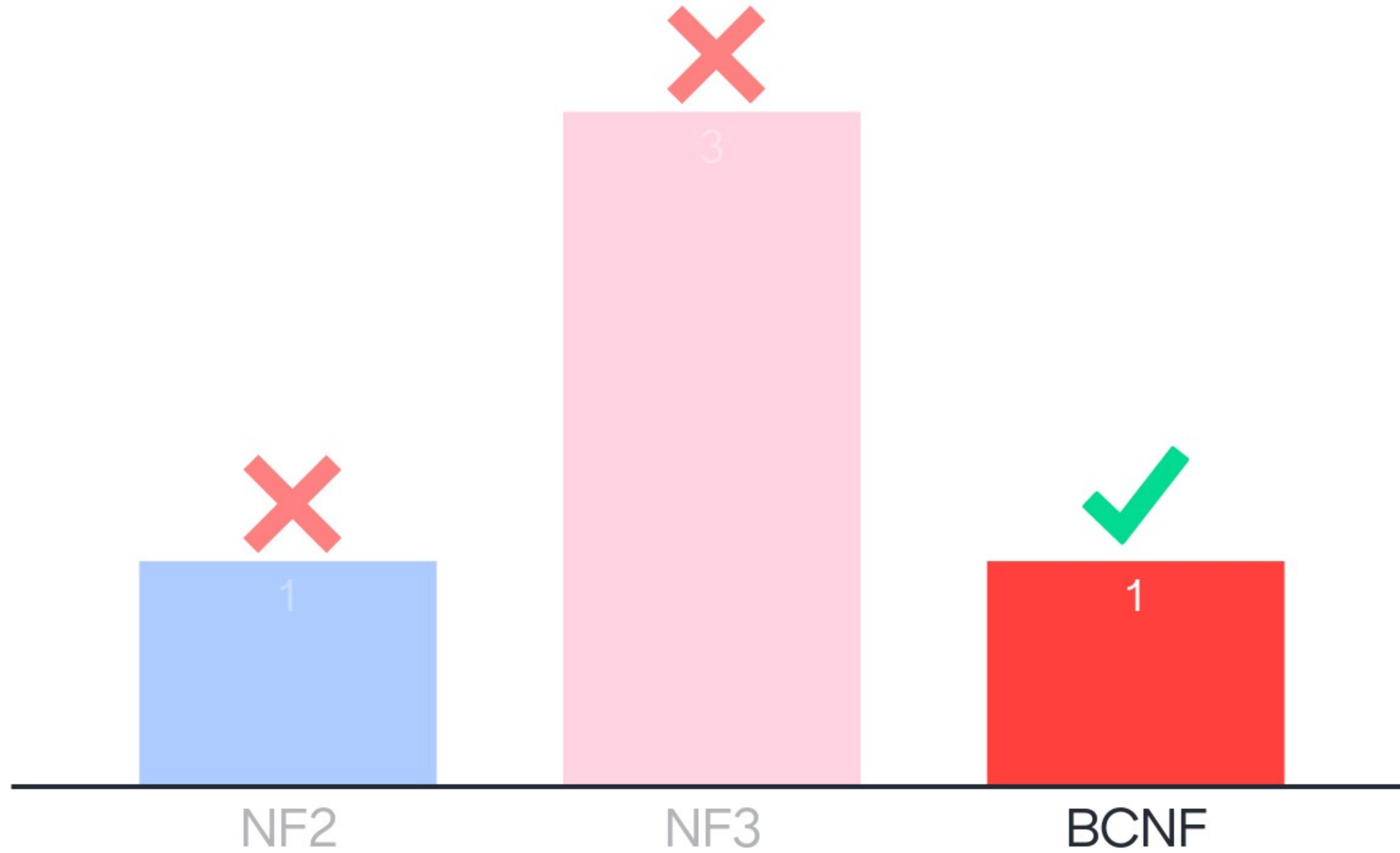
# Dette er en feil på?



# Dette er en feil på?

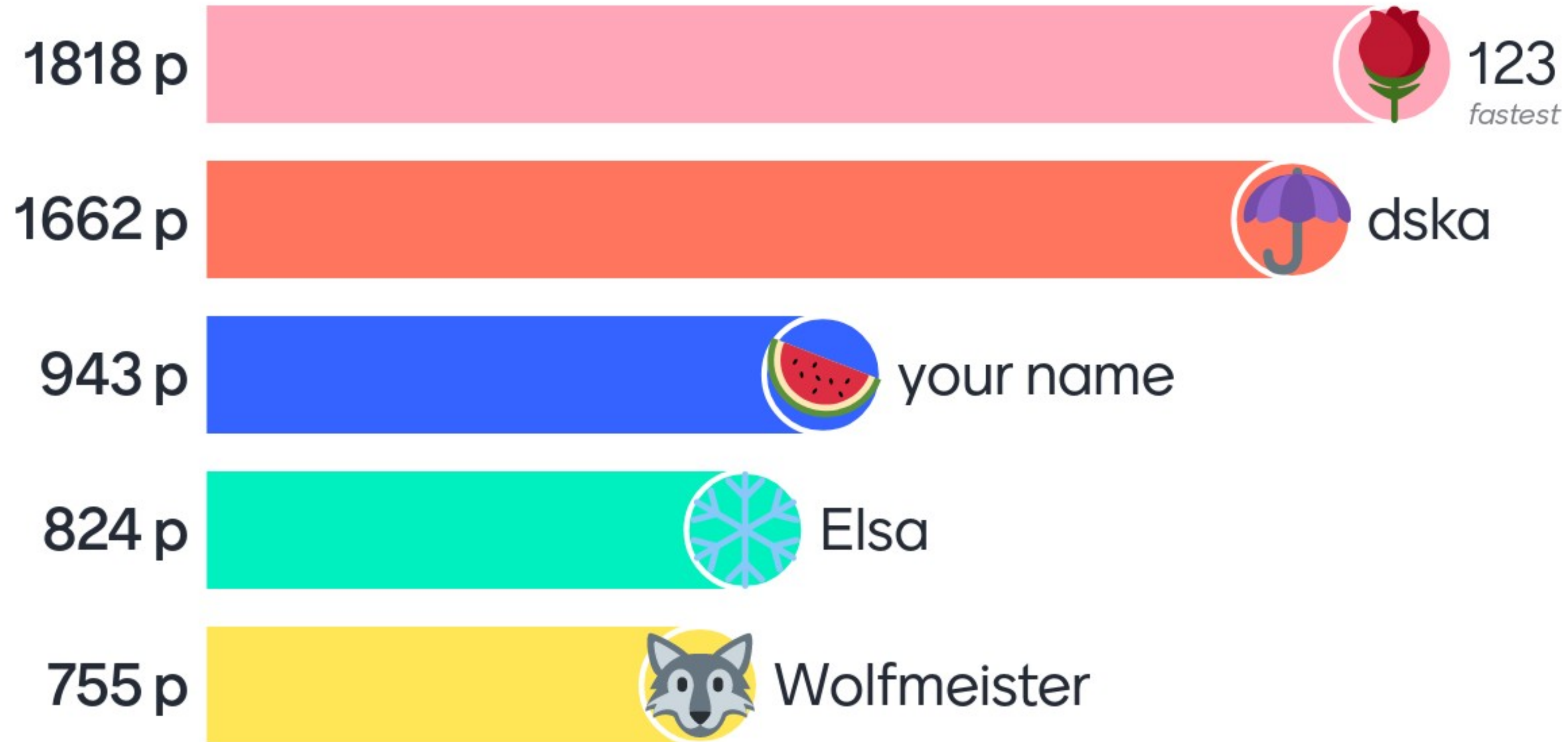


# Dette er en feil på?





# Leaderboard



# ARBETIDSTID!

=D

