Eindrapport Data Mining – Instroom Defensiepersoneel (2018–2022)

# 1. Onderzoeksvraag

“Welk krijgsmachtonderdeel heeft de hoogste instroom in de periode 2018–2022?” Doel van dit onderzoek is het analyseren van personeelsinstroom binnen Defensie en vaststellen welk onderdeel de meeste instroom heeft gekend in de afgelopen jaren. Dit is waardevol voor strategische personeelsplanning en recruitmentbeleid.

# 2. Visualisatie van de rauwe data

De oorspronkelijke dataset is afkomstig van het CBS en bevat gegevens over:  
- instroom en uitstroom van defensiepersoneel (2018–2022),  
- uitgesplitst naar geslacht, leeftijd, herkomst, en defensieonderdeel.  
  
Visualisaties zijn o.a. gegenereerd in Power BI (tijdens het vak Data Visualisatie) en ondersteund door Python.

# 3. Data-preparatie

De ruwe Excel-tabel bevatte metadata op de eerste rijen. Deze zijn verwijderd door bij het inlezen de eerste drie regels over te slaan. Vervolgens is de dataset als volgt voorbereid:  
- Gefilterd op Kenmerk == "Defensieonderdeel".  
- Verwijderd: rijen zonder instroomcijfers (Aantal == NaN).  
- Samengevoegd: totalen per krijgsmachtonderdeel over de jaren 2018–2022.

# 4. Data-analyse

Voor het beantwoorden van de onderzoeksvraag is een eenvoudige groepsanalyse toegepast:  
- Gegevens zijn gegroepeerd per krijgsmachtonderdeel (Categorie).  
- De instroomaantallen (Aantal) zijn bij elkaar opgeteld per onderdeel.  
- De totalen zijn gesorteerd van hoog naar laag.

# 5. Resultaten

De top 5 krijgsmachtonderdelen met hoogste instroom:  
1. CLAS-Manschappen – 8.900  
2. CZSK-Manschappen – 3.340  
3. BS/DOSCO/COMMIT (schaal 9+) – 2.600  
4. CLSK-Manschappen – 2.030  
5. KMAR-Manschappen – 1.880  
  
Conclusie: De Landmacht (CLAS-Manschappen) had in de periode 2018–2022 de hoogste instroom.

# 6. Machine learning technieken

Om een eenvoudige voorspelling te doen, is een lineaire regressie toegepast op de jaarlijkse instroomcijfers van CLAS (de landmacht). Het doel was om op basis van de trend van 2018 t/m 2022 een inschatting te maken van de verwachte instroom in 2023.  
  
Met behulp van de Python-bibliotheek scikit-learn is een regressiemodel getraind op de variabelen Jaar (input) en Aantal (output). Het model voorspelt een verwachte instroom van:  
  
→ ± 41.868 instromers voor CLAS in 2023.  
  
Hoewel het model slechts één variabele gebruikt (jaar) en dus relatief eenvoudig is, geeft het een eerste indicatie van een stijgende trend. In een vervolgonderzoek kunnen complexere modellen worden ingezet met meerdere variabelen zoals geslacht, leeftijdsgroep of economische factoren.

# 7. Gebruikte features

De analyse is uitgevoerd op basis van de volgende kolommen:  
- Kenmerk: om te filteren op “Defensieonderdeel”  
- Categorie: het specifieke krijgsmachtonderdeel  
- Jaar: voor eventuele tijdanalyses  
- Aantal: het daadwerkelijke instroomcijfer

# 8. Conclusie

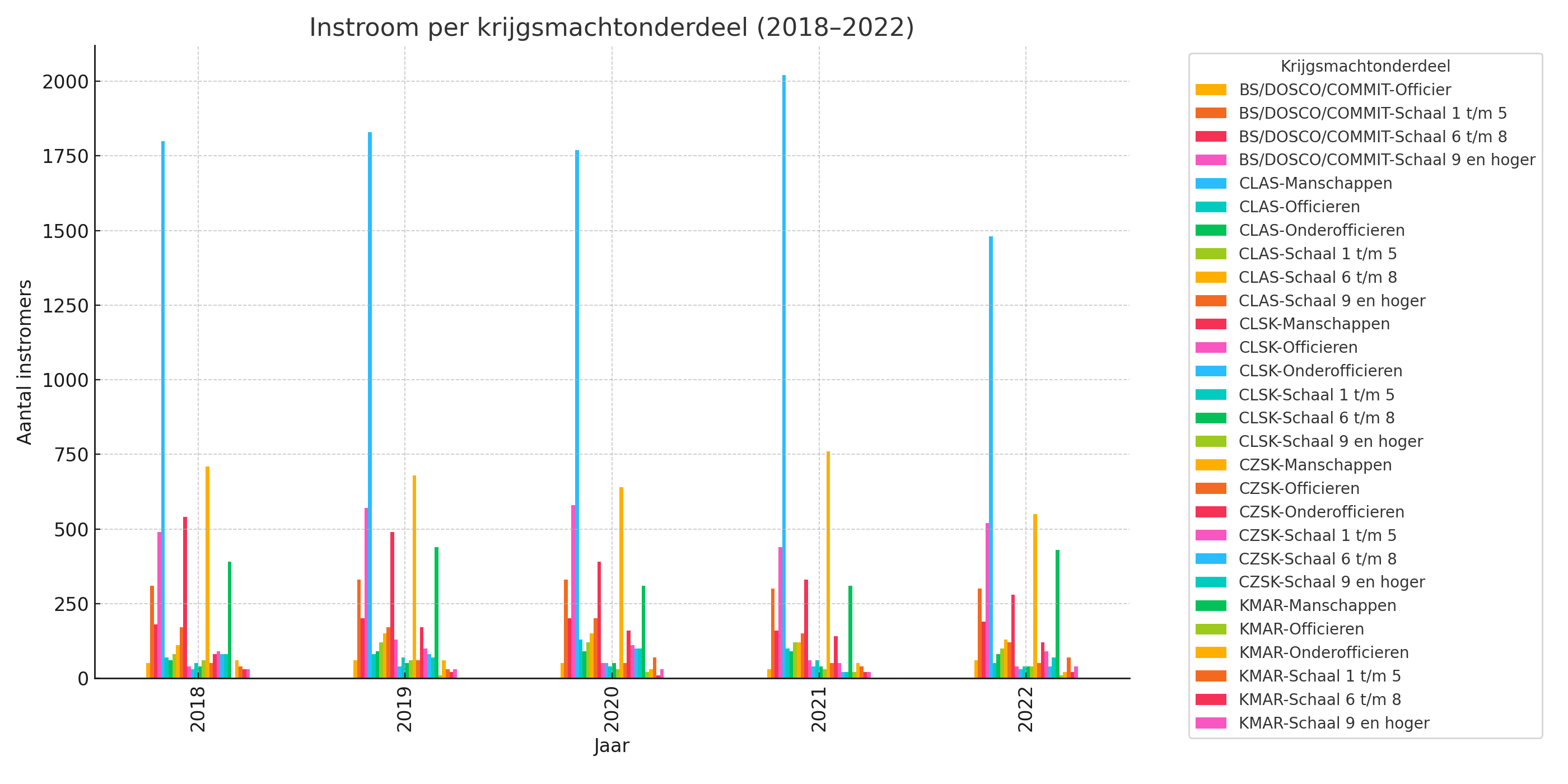
Op basis van de analyse blijkt dat CLAS-Manschappen (de Landmacht) in de periode 2018–2022 het hoogste aantal instromende personeelsleden heeft ontvangen. Dit inzicht kan gebruikt worden om de instroomverdeling over onderdelen te evalueren en het recruitmentbeleid per krijgsmachtonderdeel verder te optimaliseren.

# 📎 Bijlagen

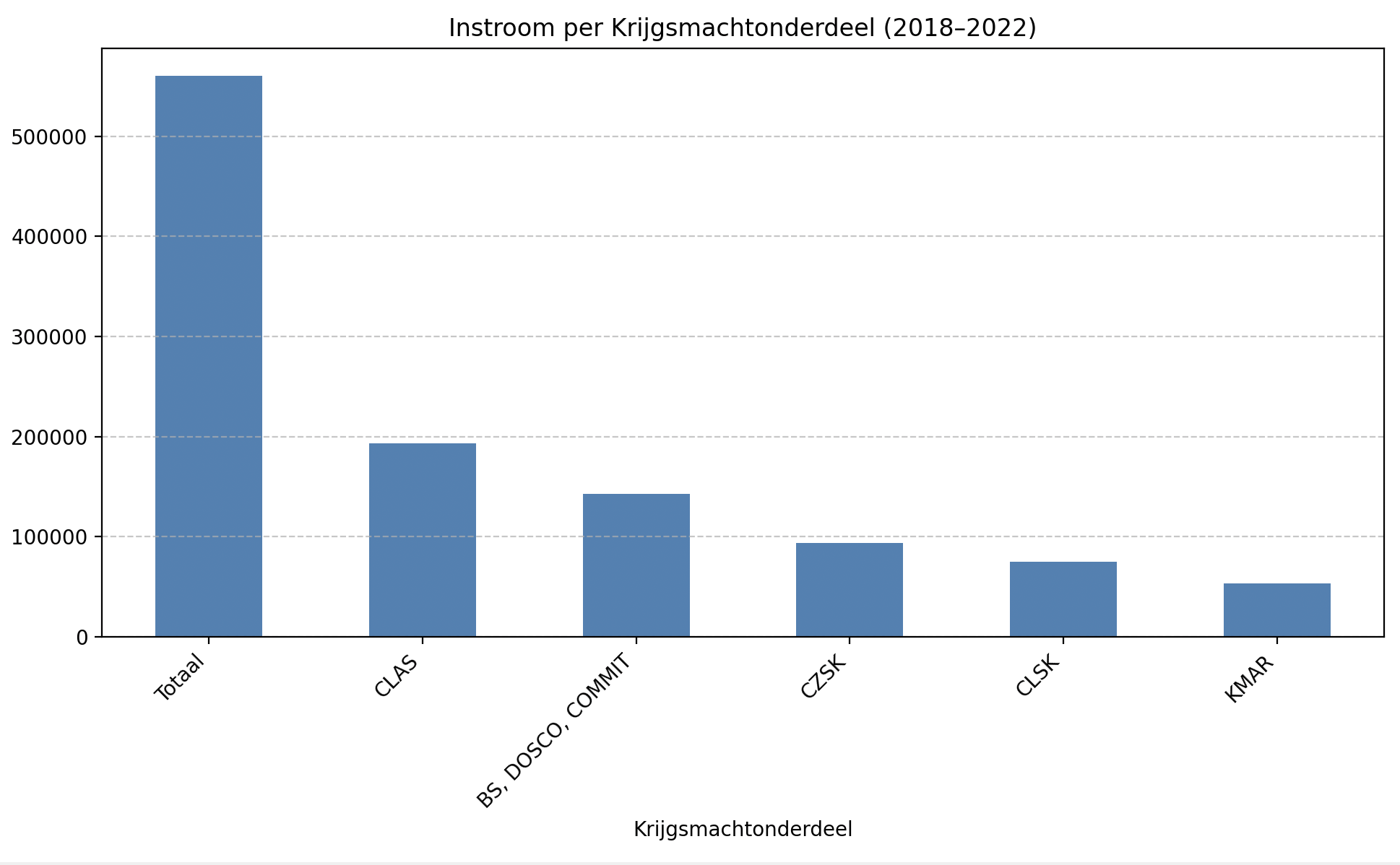
- Dataset (.csv)  
- Python-code (.py)  
- Rapport (PDF)

## Visualisatie: Instroom per krijgsmachtonderdeel (2018–2022)

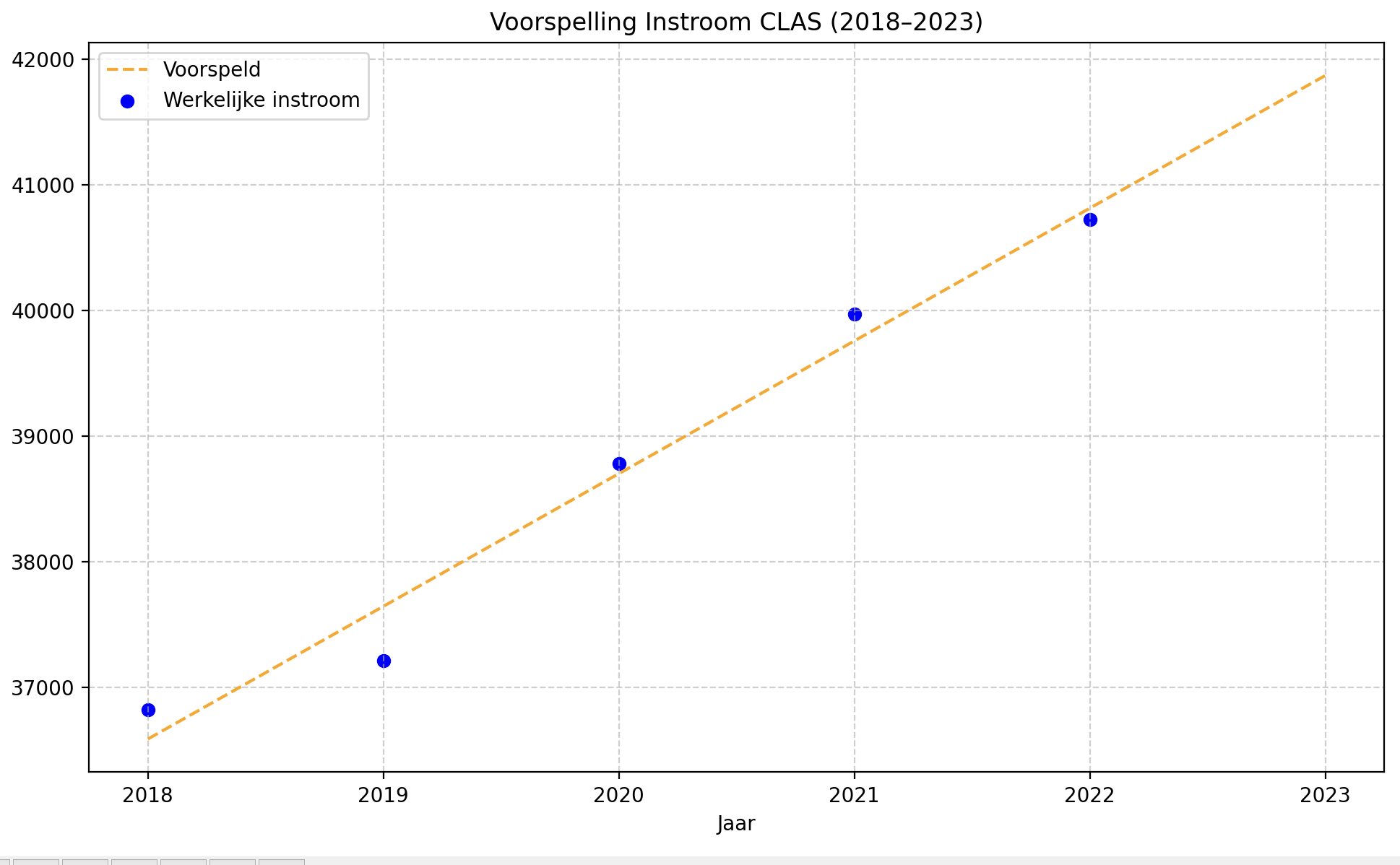
Onderstaande grafiek toont de jaarlijkse instroom per krijgsmachtonderdeel. Hierin is duidelijk zichtbaar dat CLAS-Manschappen elk jaar consistent de hoogste instroom heeft, gevolgd door CZSK en BS/DOSCO/COMMIT.



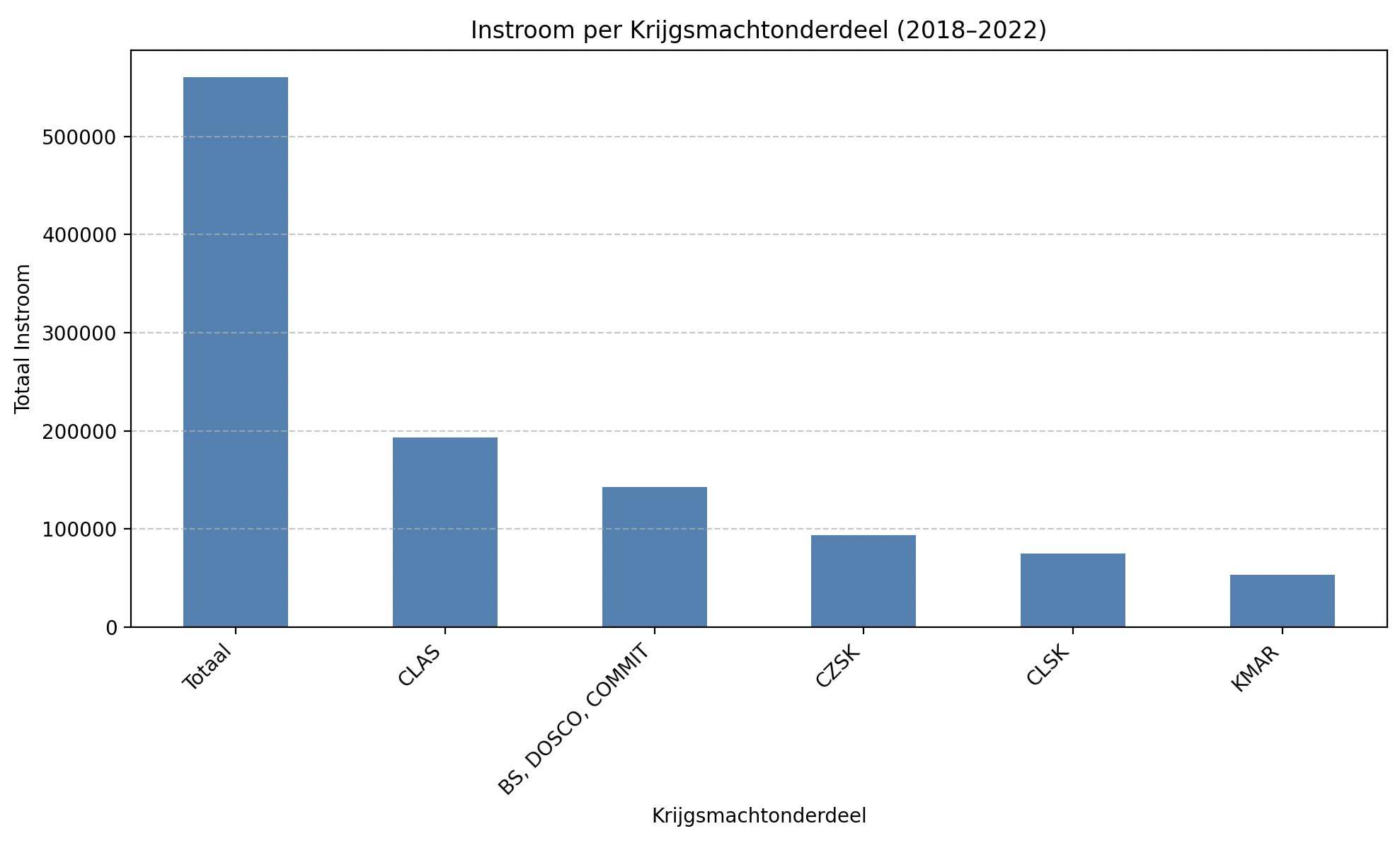
# Bijlage: Visualisaties



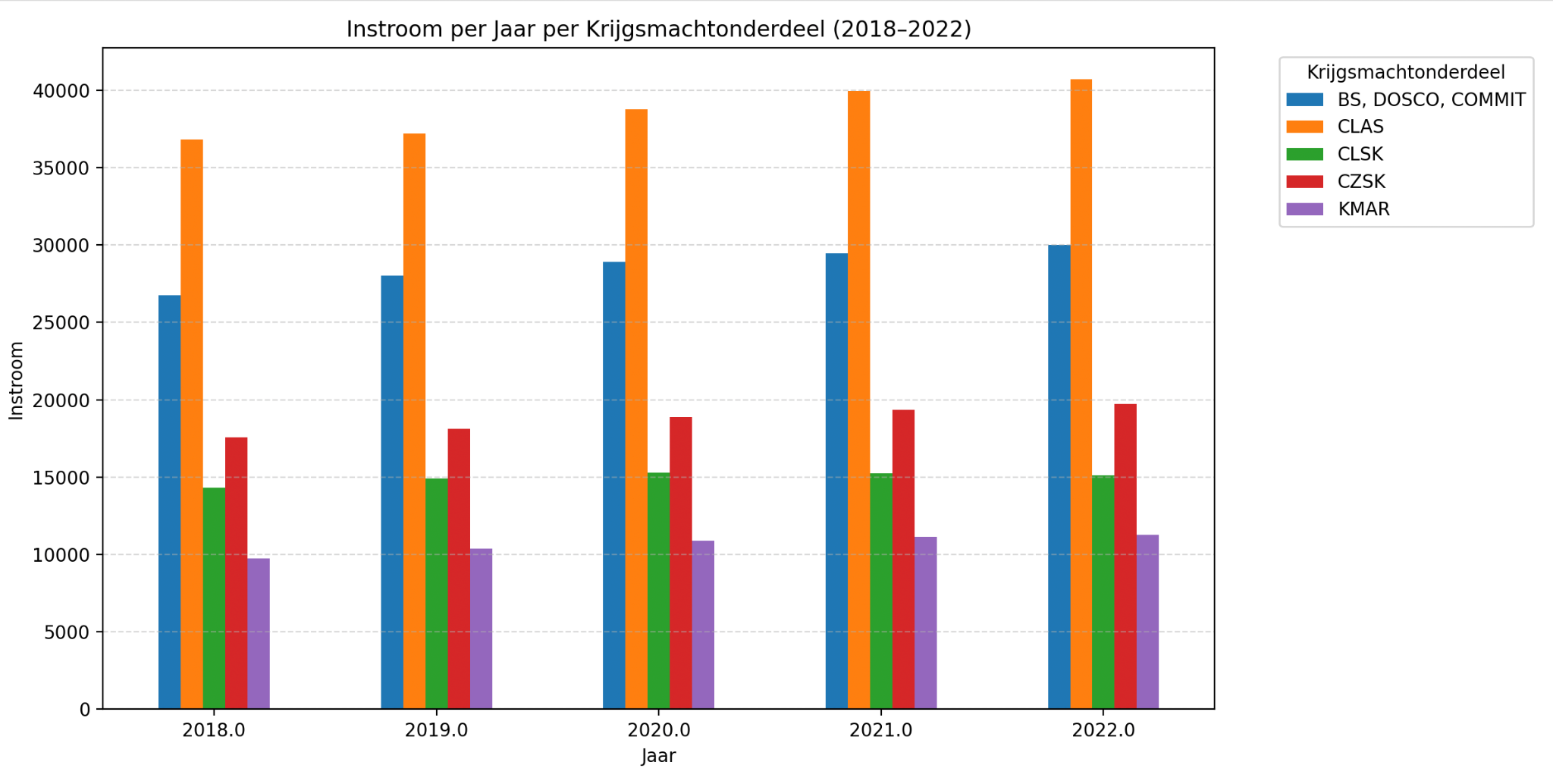
Figuur 1 – Instroom per krijgsmachtonderdeel (2018–2022)



Figuur 2 – Voorspelling instroom CLAS (2018–2023)



Figuur 3 – Instroom per jaar per krijgsmachtonderdeel



Figuur 4 – Totale instroom per krijgsmachtonderdeel (2018–2022)