

Baseline voor Noord- en Zuid Holland

Het random algoritme beschreven in deze baseline is toegepast op data voor Noord- en Zuid Holland, echter is dit te generaliseren naar heel Nederland, omdat het algoritme en de structuur van de data hetzelfde is voor beide gevallen.

Het random algoritme dat we hebben geïmplementeerd heeft de volgende specificaties.

- De lengte van elk traject wordt random bepaald, tussen 1 en de maximum tijd (120)
- Het startstation van een traject wordt random bepaald (één van de stations)
- Het aantal trajecten in het netwerk wordt random bepaald, tussen 1 en het maximum aantal (7)
-

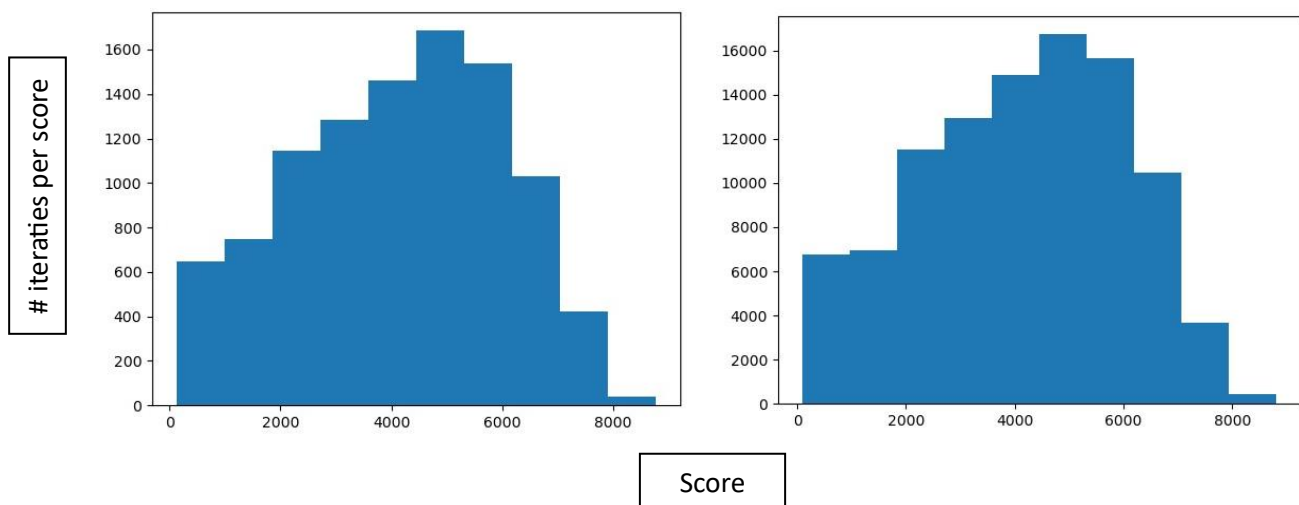
Het runnen van dit random algoritme hebben we tweemaal gedaan, één keer met 10,000 iteraties en één keer met 100,000 iteraties. Voor de run met 100,000 iteraties gaf dit een gemiddelde score van 4119.85. De score van een netwerk wordt als volgt berekend:

$$K = 10000 - (T * 100 + \text{min});$$

- K is de kwaliteit van het netwerk
- T is aantal trajecten in het netwerk
- Min is totale tijd van het netwerk

Histogram 10,000 iteraties

Histogram 100,000 iteraties



Uit de histogrammen kunnen we afleiden dat de score niet een uniforme verdeling heeft. Dit is te verklaren door te kijken naar hoe een score tot stand komt. Een netwerk bestaat uit een random aantal trajecten, en voor een uitzonderlijk hoge of lage score is er uit een laag aantal combinaties van trajecten te kiezen. Alternatief, is voor een relatief gemiddelde score de keuze uit een groot aantal combinaties van trajecten. Dus hoewel de steekproeven zelf min of meer uit een uniforme verdeling komen, is de verdeling van de score juist niet uniform.