# Thuisproject 2025 - Voorbereiding

Vul hier uw gegevens in

|  |  |
| --- | --- |
| Naam & voornaam: | Snaet Warre |

### Dataset

Kies één van de opgegeven datasets (zie Leho).

|  |  |
| --- | --- |
| Gekozen dataset: | <https://www.kaggle.com/datasets/arsri1/arrest-data-in-los-angeles/data> Arrest Data in Los Angeles |

## Bestudeer de dataset

Gebruik je kennis van data science om de volgende vragen te beantwoorden:

* Bepaal voor elke kolom het type variabele (discreet, continu, nominaal of ordinaal).
* Bereken, waar mogelijk, de centrummaten en spreidingsmaten. Wat zeggen deze maten over de data en welke inzichten leveren ze op?

## Dataset preprocessing

Geef minimum **vier** preprocessing stappen die je zou gebruiken. Geef ook op zijn minst **vier zinvolle plotjes** die je zou willen genereren en wat je er hoopt uit te leren. Let op, hou er rekening mee dat sommige waarden categorisch zijn (en bepaalde plotjes dan niet zinvol zijn).

|  |  |
| --- | --- |
|  | Beschrijving |
| Preprocessing stap 1 | Check for missing values in the set. |
| Preprocessing stap 2 | Converteren van de datum’s in de dataset naar Date Objects. |
| Preprocessing stap 3 | Checken op outliers on the age column. (geen bandieten van 150j oud bv) |
| Preprocessing stap 4 | De coördinaten omzetten naar gestandaardiseerde GeoJSON-formaten |
| Plot 1 | Aantal crimes bij mannen en vrouwen (boxplotjes van maken) Om zo te kunnen zien als er een groot verschil is wellicht in geslacht? |
| Plot 2 | Een heatmap misschien op de map met locaties waar er veel crimes gepleegd worden in Los Angeles. Hopelijk kunnen we dan identificeren van geografische "hotspots" waar arrestaties geconcentreerd zijn |
| Plot 3 | Een gestapelde staafdiagram waarbij de staven demografische groepen vertegenwoordigen (combinatie van leeftijdscategorieën en etniciteit) en de stapelingen de verschillende types arrestaties (Charge Group Description). Hopelijk kunnen we zo inzicht krijgen in hoe verschillende overtredingen verdeeld zijn over demografische segmenten |
| Plot 4 | Een violin plot die de verdeling van leeftijden toont voor elk type arrestatie (Charge Group Description), aangevuld met box plots voor statistische vergelijking. Zo kunnen we specifieke leeftijdsprofielen identificeren voor verschillende overtredingen. |

## Vier zinvolle zoekopdrachten

Geef **vier** zinvolle zoekopdrachten door. Zie projectopgave voor meer context.

* Bij minimum drie opdrachten moeten er vooraf één of meerdere parameters opgevraagd worden om zo de zoekopdracht te verfijnen.
* Bij minimum één zoekopdracht moet er een grafiek kunnen gegenereerd worden.

|  |  |
| --- | --- |
|  | Beschrijving zoekopdracht |
| Zoekopdracht 1 | **1. Arrestaties per Gebied en Tijdsperiode** **Parameters:**   * Gebied (Area Name): dropdown met alle beschikbare gebieden * Startdatum (Start Date): datumkiezer * Einddatum (End Date): datumkiezer * Leeftijdsbereik (Age Range): slider of min/max velden |
| Zoekopdracht 2 | **2. Trend van Specifieke Overtreding over Tijd** **Parameters:**   * Arrestatietype (Charge Group Description): dropdown met alle beschikbare overtredingstypen * Tijdsgranulariteit: dropdown (dagelijks, wekelijks, maandelijks, jaarlijks) * Gebied (optioneel): multi-select dropdown voor specifieke gebieden |
| Zoekopdracht 3 | **3. Demografische Analyse van Arrestaties** **Parameters:**   * Geslacht (Sex Code): multi-select (M/F) * Etniciteit (Descent Code): multi-select met alle beschikbare etniciteitscodes * Arrestatietype (Charge Group Description): dropdown (optioneel) |
| Zoekopdracht 4 | **4. Geografische Hotspots van Arrestaties** **Parameters:**   * Radius (in km): numeriek invoerveld * Centrumpunt (LAT/LON): kaart waarop een punt kan worden geselecteerd of adresinvoer * Tijdsbereik: datumkiezer voor begin- en einddatum * Arrestatietype (optioneel): multi-select met arrestatietypes |