# Praxisprojekt Datenvisualisierung

## Aleksej Assmus, Laura Richter, Sarah Leibach

### Datensatz:

[https://www.kaggle.com/datasets/frtgnn/dog-breeds-by-personality-and-size#](https://www.kaggle.com/datasets/frtgnn/dog-breeds-by-personality-and-size)

Vergleich der verschiedenen Hunderassen

### Extra Ideen:

* Statistiken runterladen
* empfohlene Fütterungsart
* gesundheitliche Einschränkungen

### Minimal- und Maximallösung:

**1. MUST HAVE (Grundvoraussetzungen für Bestehen)**

* **Datenexploration**
  + Grundlegende Statistiken (Mittelwerte, Korrelationen, Häufigkeiten) zu Hunderassen, Persönlichkeitsmerkmalen und Größe.
  + Prüfung auf fehlende/inkonsistente Daten.
* **10 informative Visualisierungen** (z. B.):
  + Verteilung der Hunderassen nach Größe (Small/Medium/Large).
  + Zusammenhang zwischen Persönlichkeitsmerkmalen (z. B. "Energy Level") und Größe.
  + Top 5 Rassen pro Merkmal (z. B. "Friendliest", "Most Trainable").
  + Heatmap/Korrelationsmatrix der Persönlichkeitsmerkmale.
* **Python-Code** (Jupyter Notebook) mit:
  + Datenbereinigung, Analyse (Pandas, NumPy) und Visualisierung (Matplotlib/Seaborn/Plotly).
* **Bericht/Dashboard** (Notebook oder PPT) mit:
  + Kurzer Beschreibung der Methodik, Ergebnissen und Grafiken.
* **Präsentation (15 Min.)**
  + Klare Struktur: Problemstellung, Daten, Erkenntnisse, Visualisierungen.

**2. SHOULD HAVE (Wichtige Verbesserungen)**

* **Externe Statistiken ergänzen** (z. B.):
  + Beliebtheit der Rassen (Quellen wie AKC, nationale Hundeverbände).
  + Durchschnittliche Lebenserwartung pro Rasse.
* **Interaktive Visualisierungen** (z. B. Plotly-Dashboards).
* **Professionelles Design** der Grafiken (Farben, Labels, Titel).

**3. COULD HAVE (Optional, wenn Zeit bleibt)**

* **Machine Learning Ansatz** (z. B. Clustering von Rassen nach Persönlichkeit).
* **Vergleich mit anderen Datensätzen** (z. B. Gesundheitsdaten zu Rassen).
* **Kreative Visualisierungen** (z. B. "Hundeprofile" mit Emojis/Symbolen).

**4. WON’T HAVE (Keine Priorität)**

* + Tiefgehende statistische Tests (z. B. Hypothesentests).
  + Eigenes Datenscraping oder umfangreiche Datenaufbereitung.

### Projektaufteilung:

Tägliche Rotation in den Bereichen *Bericht*, *Analyse* und *Visualisierung*, sodass jede/r alles einmal gemacht hat