



## Anfertigen eines CAT

### Data Science Kernaussage:

Durch explorative Datenanalyse (EDA) können Notrufdaten systematisch ausgewertet werden, um Muster in Zeit, Ort und Art der Notrufe zu erkennen und diese visuell aufzubereiten. Diese Erkenntnisse unterstützen eine fundierte Entscheidungsfindung im öffentlichen Sicherheitsmanagement.

### (P) roblem:

Welche Notrufarten treten wann und wo besonders häufig auf, und lassen sich Muster erkennen, die bei der Planung von Rettungseinsätzen helfen?

### (I) ntervention:

Verwendung der Bibliotheken:

- pandas zur Datenvorverarbeitung und Analyse
  - matplotlib & seaborn zur Visualisierung
- Konkret:
- Gruppierung der Einsätze nach „Reason“, „Wochentag“, „Stunde“, „Datum“
  - Erstellen von Zeitreihen und Heatmaps

### (K) ontrollintervention

Einsatz anderer Analysewerkzeuge oder -ansätze, z. B.:

- Prädiktive Modelle mit scikit-learn
- Geovisualisierung mit folium Diese wurden im aktuellen Projekt **nicht** verwendet, wären aber weiterführende Möglichkeiten.

### (E) rgebnismaß (Zielgröße(n)) – Die Evidence

- Übersicht über Einsatzgründe (z. B. Fire, EMS, Traffic)
- Identifikation zeitlicher Muster (Tageszeit, Wochentag, Monat)
- Visualisierung der Daten zur besseren Interpretation und Entscheidungsunterstützung
- Grundlage für zukünftige Planungen im Rettungswesen

### Anmerkungen / Literaturhinweise

- Kaggle Dataset: 911 Calls Dataset
- Wes McKinney: *Python for Data Analysis*
- Seaborn & Pandas Dokumentationen
- educX Lernmaterialien

## Die Suche nach der besten Evidenz

### 1. Problem

#### Definition:

Ermittlung von Mustern in Notrufdaten, um herauszufinden, welche Anrufgründe (z. B. EMS, Fire, Traffic) in welchen Zeiträumen (Stunden, Wochentagen, Monaten) besonders häufig vorkommen.

#### Kundennutzen:

Diese Information kann dazu beitragen, Ressourcen und Einsatzkräfte im öffentlichen Sicherheitsbereich gezielt zu planen und Engpässe zu vermeiden.

---

### 2. Definition einer wichtigen suchbaren Frage

#### Beispielfrage:

„Welcher Notrufgrund tritt am häufigsten auf und wie variieren diese Einsätze in Abhängigkeit von Tageszeit, Wochentag und Monat?“

#### Ziel:

Die Frage lässt sich durch den Einsatz von Gruppenanalysen und Visualisierungstechniken im Python-Code beantworten, indem Du die Daten aufbereitet und aggregiert hast.

---

### 3. Auswahl der wahrscheinlichsten Quelle für diese Evidenz

#### Primäre Quelle:

Der bereitgestellte 911-Notrufdatensatz von Kaggle.

#### Begründung:

Dieser Datensatz enthält die relevanten Felder (u. a. Titel, Zeitstempel, Ort) und ist bereits für eine explorative Analyse vorgesehen – er bietet eine solide Basis, um Trends und Muster zu erkennen.

---

### 4. Erstellung einer Suchstrategie

#### Ansatz:

- **Datenimport und -aufbereitung:** Einlesen des CSV-Datensatzes in ein Pandas DataFrame und Umwandeln der Zeitstempel in das DateTime-Format.
- **Feature Engineering:** Erzeugen von zusätzlichen Spalten (z. B. „Reason“, „Hour“, „Month“, „Day of Week“) mittels `.apply()` und `.map()`, um die relevanten Informationen aus den vorhandenen Daten zu extrahieren.
- **Explorative Analysen:**

- Zählen der häufigsten Notrufgründe (mittels `value_counts()` und `countplot`).
  - Gruppierung der Daten nach Zeitkomponenten (z. B. Gruppierung nach Monat oder Wochentag) und Erstellung von Visualisierungen (Liniendiagramme, Heatmaps, Clustermaps).
  - **Dokumentation:** Sorgfältige Aufzeichnung aller Analyse-Schritte, um die Ergebnisse nachvollziehbar zu machen.
- 

## 5.0 Zusammenstellung der Evidenzausbeute

### Evidenzausbeute:

- **Resultate aus den Analysen:**
    - Häufigste Notrufgründe werden durch die extrahierte „Reason“-Spalte identifiziert (z. B. EMS dominiert die Aufrufzahlen).
    - Zeitliche Muster: Analyse der Verteilung der Notrufe über Stunden, Tage und Monate zeigt, wann und wie sich Einsätze häufen.
    - Visualisierungen (Countplots, Heatmaps, Clustermaps) bieten eine anschauliche Darstellung der Einsatzmuster und unterstützen die Interpretation der Daten.
- 

## 5.1 Falls die Evidenzausbeute schlecht ist

### Alternative Strategien:

- **Wahl einer zweitwahrscheinlichen Quelle:**

Falls sich herausstellt, dass der Kaggle-Datensatz unvollständige oder ungenaue Informationen liefert, könnte man zusätzliche Datenquellen (z. B. weitere öffentliche Datensätze oder ergänzende Verwaltungsdaten der Notrufstellen) heranziehen.
  - **Erstellung einer überarbeiteten Suchstrategie:**
    - Erweiterung der Analyse um zusätzliche Variablen, z. B. geographische Clusteranalysen oder Einbezug von Wetterdaten.
    - Einsatz anderer Bibliotheken (z. B. `scikit-learn` für mögliche Klassifikationsaufgaben oder prädiktive Modelle) zur Validierung der Ergebnisse.
  - **Zusammenfassung und kritische Bewertung der evidenzbasierten Erkenntnisse:**
    - Dokumentation der Limitationen der ursprünglichen Analyse und Darstellung der Vorteile der alternativen Ansätze.
- 

## 6. Anwendung der Evidenz

### Ziel der Anwendung:

- **Verbesserte Einsatzplanung:**

Die gewonnenen Erkenntnisse sollen Entscheidungsträgern helfen, Ressourcen im

Rettungsdienst gezielt einzusetzen (z. B. durch Erhöhung der Personalstärke in identifizierten Spitzenzeiten).

- **Basis für weitere Analysen:**

Die evidenzbasierte Darstellung der Einsatzmuster kann als Grundlage dienen, um prädiktive Modelle zu entwickeln oder gezielte Maßnahmen zur Verbesserung der Notfallreaktionen abzuleiten.

- **Kommunikation der Ergebnisse:**

Die visualisierten Daten (Diagramme, Heatmaps) erleichtern es, komplexe Trends verständlich an das Management oder externe Partner zu kommunizieren und unterstützen so die Implementierung von Handlungsempfehlungen.