**#README Datenfile `hackathon.csv`**

Der Datensatz `hackathon.csv` enthält anonymisierte Informationen zu den zurückgelegten Wegstrecken der Teilnehmer von track4science (2024-03-24 bis 2024-09-13).

* Der Datensatz enthält nur Informationen zu einzelnen Wegstrecken. Eine Wegstrecke ist definiert als einen Streckenabschnitt mit einem einheitlichen Verkehrsmittel.
* Beschrieben wird die Wegstrecke anhand von
  + ***ID*:** Indikator für die Kombination aus Person und Datum der Wegstrecke.
  + ***StartedAt\_Timestamp*** und ***FinishedAt\_Timestamp***: Start- und Endzeiten jeder Wegstrecke.
  + ***Length*:** Die zurückgelegte Strecke in m.
  + ***Mode:*** Das verwendete Verkehrsmittel (z. B. zu Fuß, Fahrrad, etc.).
  + ***Geometry3\_JSON:***Sequenz von GPS-Koordinaten einer Wegstrecke im JSON-Format:
    - *coordinates****:*** Hauptobjekt, das eine Sammlung von Koordinatenpaaren enthält.
    - *V1, V2…:* Aus Datenschutzgründen wurden bestimmte Koordinaten, die sensible Bereiche betreffen (wie Wohngebiete oder andere private Orte), entfernt. Die V-Einträge markieren die ursprüngliche Position der Koordinatenpaare in der Reihenfolge des Originalpfades, um Lücken oder ausgelassene Punkte zu kennzeichnen. Beispielsweise bedeutet ein Eintrag wie V3, dass dieses Koordinatenpaar ursprünglich das dritte Paar in der vollständigen Datenreihe war.
    - *Format der Koordinatenpaare:* Die Einträge enthalten Werte im WGS84-Koordinatensystem, das Längengrad und Breitengrad in Dezimalgrad angibt, z. B. [9.2244, 49.129].
    - Sofern alle GPS-Koordinaten entfernt wurden, wird eine leere geschweifte Klammer angegeben {}.
    - *type:* Ein Attribut mit dem Wert "LineString", das angibt, dass die Koordinatenpaare eine fortlaufende Linie darstellen.
* Der Datensatz enthält eine zufällige Auswahl von 1000 Personen/Datum-Kombinationen. Er beinhaltet für diese Kombination alle Wegstrecken, die eine Person zu diesem Datum getätigt hat und die in einem Radius von 5 km um den Bildungscampus Heilbronn gestartet oder beendet wurden (basierend auf dem ersten oder letzten GPS-Koordinatenpaar einer Wegstrecke). Die ID gibt die Personen/Datum-Kombination an. Einzelne Personen können in dem Datensatz mehrfach enthalten sein.
* Innerhalb einzelner Wegstrecken wurden einzelne GPS-Koordinatenpaare entfernt, sodass die dargestellten GPS-Koordinaten auch nur Teile einer Wegstrecke abdecken:
  + Sämtliche GPS-Koordinaten außerhalb eines Radius von 5 km um den Bildungscampus Heilbronn wurden entfernt.
  + Um den exakten Wohnort zu verschleiern, wurden sämtliche GPS-Koordinaten innerhalb eines Radius von 500m um das erste und letzte GPS-Koordinatenpaar eines Tages entfernt.
  + Um sensible Aufenthaltsorte zu verschleiern, wurden sämtliche GPS-Koordinaten innerhalb eines Radius von 500m um einen Aufenthaltsort entfernt, sofern der Aufenthalt länger als 30 Minuten gedauert hat. Umstiege in ein anderes Verkehrsmittel mit einer Umsteigezeit von weniger als 30 Minuten sind in den Daten vollständig abgebildet.
* Um die Privatsphäre der Nutzer zu schützen, wurden die GPS-Koordinaten auf ein Raster (100m \* 100m) projiziert und jedes Koordinatenpaar auf den Mittelpunkt des Rasters verschoben. Diese Anpassung bewahrt den Kontext der Bewegungsmuster, anonymisiert jedoch die exakten Standorte, um eine Verknüpfung mit anderen GPS-Datensätzen zu erschweren.
* Das Datum der Wegstrecke wurde für alle Wegstrecken verfremdet und uniform auf 2024-11-18 gesetzt. Für alle Personen/Datumskombinationen wurde der Start- und Endzeitpunkt um eine zufällige Zeitdauer (+/- 60 Minuten) verschoben. Damit sind keine genauen Zeitpunkte zurückverfolgbar, das Bewegungsmuster und die zeitliche Struktur bleiben jedoch erhalten.