

Dokumentation WetterStationen

Allgemeine Kriterien zur Bewertung der Datenqualität:

1. Vollständigkeit: Sind alle erwarteten Daten vorhanden oder fehlen einige Monate/Wetterstationen?
2. Konsistenz: Sind die Daten logisch miteinander verknüpft? Gibt es keine Widersprüche in den Daten, wie zum Beispiel negative Niederschlagswerte?
3. Genauigkeit: Sind die Daten korrekt und repräsentativ für den tatsächlichen Niederschlag in den Wetterstationen?
4. Aktualität: Sind die Daten auf dem neuesten Stand oder sind sie veraltet?

Maßnahmen zur Verbesserung der Datenqualität:

1. Überprüfung der Daten auf Vollständigkeit und Korrektheit.
2. Korrektur von offensichtlichen Fehlern oder Ausreißern.
3. Ergänzung fehlender Daten, falls möglich.
4. Überprüfung der Daten auf Plausibilität und Konsistenz.
5. Standardisierung der Daten, um Vergleiche zwischen verschiedenen Wetterstationen zu ermöglichen.

Methoden zur Reduzierung der Datenredundanz:

1. Zusammenfassung der Daten auf eine höhere Stufe, z.B. durch Berechnung von Jahres- oder Quartalsdurchschnittswerten anstatt Monatswerten.
2. Auswahl von repräsentativen Wetterstationen, um die Datenmenge zu reduzieren.
3. Entfernung von unwichtigen oder redundanten Variablen, um den Speicherbedarf zu reduzieren.
4. Verwendung von Datenkompressionstechniken, um den Speicherbedarf zu reduzieren.
5. Aggregation von Daten über bestimmte geographische Regionen, um die Datenmenge zu reduzieren.

Dokumentation der Datenqualitätserhöhung

Im Rahmen der Erhöhung der Datenqualität wurden folgende Schritte vorgenommen. Die Originaldatei blieb unberührt von den Veränderungen, stattdessen wurden die Dateien in eine neue Datei transponiert.

1. Formatierung der Monatsspalten: Die Monatsspalten waren unterschiedlich formatiert, was die Analyse erschwert hätte. Deshalb wurden die Monatsspalten einheitlich formatiert, indem alle Werte in den Spalten als Dezimalzahl angegeben wurde und auf eine Nachkommastelle gerundet.
2. Umwandlung der Jahre in Jahreszahlen: In der Spalte "Jahr" waren die Jahre als vierstellige Zahlen dargestellt. Um die Lesbarkeit zu verbessern und die weitere Bearbeitung zu erleichtern, wurden die Zahlen in tatsächliche Jahreszahlen umgewandelt. Hierfür wurde in Excel die Formel "`=Jahre(X)`" genutzt, wobei X für die jeweilige Zelle steht, in der das Jahr steht.
3. Überprüfung auf fehlende oder inkonsistente Daten: Es wurde überprüft, ob es fehlende Daten oder inkonsistente Angaben gibt. Fehlende Daten können beispielsweise durch eine fehlerhafte Messung entstanden sein oder durch einen Datenverlust. Inkonsistente Angaben können durch einen Eingabefehler oder durch eine fehlerhafte Messung entstanden sein.
4. Entfernung von Duplikaten: Duplikate können durch verschiedene Faktoren entstehen, beispielsweise durch eine Mehrfachmessung an der gleichen Station oder durch einen Fehler bei der Dateneingabe. Duplikate der Wetterstationen wurden entfernt, um die Analyse nicht zu verfälschen.
5. Bereinigung von fehlerhaften Daten: Fehlerhafte Daten können durch verschiedene Faktoren entstehen, beispielsweise durch eine fehlerhafte Messung oder durch eine fehlerhafte Eingabe. Fehlerhafte Daten wurden bereinigt, um eine korrekte Ausarbeitung zu ermöglichen.

Verbesserungsvorschläge für zukünftige Projekte mit Daten:

1. Dokumentation der Datenerhebung und -verarbeitung: Um eine korrekte Analyse zu ermöglichen, sollten alle Schritte der Datenerhebung und -verarbeitung dokumentiert werden. Hierdurch kann nachvollzogen werden, welche Schritte vorgenommen wurden und welche Daten zur Verfügung stehen.
 2. Verwendung standardisierter Formate: Um die Analyse zu erleichtern, sollten standardisierte Formate für die Datendarstellung verwendet werden. Hierdurch können Daten einfacher verglichen und analysiert werden.
 3. Regelmäßige Überprüfung der Datenqualität: Um eine korrekte Analyse zu gewährleisten, sollten die Daten regelmäßig auf Qualität überprüft werden. Hierdurch können fehlerhafte Daten frühzeitig erkannt und bereinigt werden.
-

Technische Dokumentation

Der Code liest eine CSV-Datei ein und zeigt deren Inhalte in einem Treeview-Widget an. Der Nutzer kann die Tabelle durch Klicken auf die Spaltenüberschriften sortieren lassen. Zusätzlich gibt es einen Button, mit dem die Tabelle nach der Summe der Niederschlagsmenge jedes Jahres sortiert werden kann.

Zu Beginn wird die Pandas-Bibliothek importiert, um die CSV-Datei zu lesen und zu verarbeiten, sowie die Tkinter-Bibliothek, um das GUI-Fenster und die Widgets zu erstellen.

Die CSV-Datei wird mit der Funktion `pd.read_csv()` eingelesen und in einem Pandas-Dataframe gespeichert. Dabei wird der Pfad zur Datei angegeben, der Trennzeichen- und Dezimalpunkttyp definiert sowie Duplikate in der Spalte "Stations_id" entfernt.

Das GUI-Fenster wird mit der Funktion `Tk()` von Tkinter erstellt und die Größe auf 1400×500 Pixel festgelegt.

Das Treeview-Widget wird mit der Funktion `ttk.Treeview()` von Tkinter erstellt. Die Spaltenüberschriften werden in das Widget eingefügt und mit der Funktion `SortColumn()` sortiert, wenn sie angeklickt werden.

Die Funktion `SortColumn()` erhält als Argument die Spaltenüberschrift, nach der sortiert werden soll, sowie optional die Sortierreihenfolge. Die Funktion löscht zunächst die aktuelle Ansicht im Treeview-Widget und sortiert die Daten im dataframe nach der ausgewählten Spalte. Danach wird die sortierte Tabelle in das Treeview-Widget eingefügt.

Die Funktion `SortByNiederschlagsmenge()` wird aufgerufen, wenn der Button "Nach Jahresniederschlagsmenge sortieren" angeklickt wird. Die Funktion berechnet die Summe der Niederschlagsmenge für jedes Jahr im Datenrahmen, sortiert die Tabelle nach der berechneten Summe und entfernt die zusätzliche Spalte mit den Summen. Danach wird die sortierte Tabelle in das Treeview-Widget eingefügt.

Zum Schluss wird das Treeview-Widget mit den Daten und den Spaltenüberschriften gefüllt und in das GUI-Fenster eingefügt. Der Button "Nach Jahresniederschlagsmenge sortieren" wird erstellt und das Fenster mit `mainloop()` geöffnet, um auf Benutzereingaben zu warten.

Kundendokumentation: Wetterdaten-Viewer

Willkommen beim Wetterdaten-Viewer! Mit diesem Tool können Sie Wetterdaten aus einer CSV-Datei visualisieren und sortieren.

Systemanforderungen

Um den Wetterdaten-Viewer nutzen zu können, benötigen Sie Folgendes:

- Python 3.x
- Die Python-Module pandas und tkinter

Installation

1. Laden Sie den Code von GitHub herunter und entpacken Sie das ZIP-Archiv.
2. Öffnen Sie die Datei `WetterdatenViewer.py` in einem Python-Editor oder einer Python-IDE.
3. Speichern Sie die Datei ab.
4. Öffnen Sie eine Konsole oder ein Terminal und wechseln Sie in das Verzeichnis, in dem die Datei `WetterdatenViewer.py` gespeichert ist.
5. Geben Sie den Befehl `python WetterdatenViewer.py` ein, um das Tool zu starten.

Verwendung

Schritt 1: CSV-Datei laden

Nach dem Start des Tools wird automatisch die Datei `Niederschlag_Daten_ungefiltert.csv` aus dem entsprechenden Ordner geladen. Wenn Sie eine andere CSV-Datei verwenden möchten, ändern Sie den Dateipfad in Zeile 6 des Codes entsprechend ab.

Schritt 2: Daten anzeigen

Nachdem die Daten geladen wurden, werden sie in einer Tabelle angezeigt. Die Spaltenüberschriften dienen als Sortierbuttons. Klicken Sie auf eine Spaltenüberschrift, um die Daten nach dieser Spalte zu sortieren.

Schritt 3: Daten filtern

Um die Daten nach Jahresniederschlagsmenge zu filtern, klicken Sie auf den Button "Nach Jahresniederschlagsmenge sortieren". Die Daten werden dann nach der Summe der Niederschlagsmengen für jedes Jahr sortiert.

Schritt 4: Sortierreihenfolge ändern

Wenn Sie die Sortierreihenfolge ändern möchten, klicken Sie einfach erneut auf den entsprechenden Sortierbutton. Die Sortierreihenfolge wird dann umgekehrt.

Hinweise

- Wenn Sie die Sortierreihenfolge ändern, wird die Filterung nach Jahresniederschlagsmenge zurückgesetzt.
 - Wenn Sie eine andere CSV-Datei laden möchten, muss diese die gleiche Struktur wie `Niederschlag_Daten_ungefiltert.csv` haben.
 - Die Spaltenbreite ist auf 80 Pixel festgelegt und kann manuell geändert werden.
 - Wenn es mehr Daten als Platz in der Tabelle gibt, können Sie die Scrollfunktion benutzen, um durch die Tabelle zu navigieren.
-