

Project Description: CSV Data Cleaner

Overview:

Das Hauptziel dieses Projekts ist die Erstellung einer robusten und vielseitigen Klasse namens **CleanerCSV**. Diese Klasse sollte in der Lage sein, CSV-Datensätze zu verarbeiten und dabei mehrere gängige Datenreinigungsaufgaben zu integrieren.

1. Datenlesen:

- Eine Methode, die ausschließlich dem Lesen eines CSV-Datensatzes gewidmet ist, sowohl in seiner Originalversion als auch in einer Kopie für die Bearbeitung.

2. Umfassende Datenreinigung:

- Das Herzstück der Klasse, bestehend aus einer Reihe von Methoden zur Behandlung von:
 - Entfernen von doppelten Zeilen.
 - Ausschluss von Zeilen mit NaN-Werten.
 - Entfernen von Spalten basierend auf der Spezifikation.
 - Leerzeichen in bestimmten Zeichenketten-Spalten abschneiden.
 - Entfernen von Zeilen, die mit den Kopfzeilen identisch sind.
 - Korrektur von Datentypen für angegebene Spalten.

3. Datenexport:

- Ein Hilfsprogramm zum Speichern des bereinigten Datensatzes in einer angegebenen CSV-Datei.

4. Datenprofiling:

- Durch Integration des Moduls **ydata_profiling** sollte die Klasse in der Lage sein, einen detaillierten Bericht zu erstellen, der Einblicke, Verteilungen und potenzielle Probleme sowohl für die Original- als auch für die bereinigten Daten zeigt.

Lieferungen:

Meilenstein: Mittwoch 28.02.2024 um 16:00

- Der fertige **CleanerCSV**-Klassen-Code.
- Eine umfassende Bedienungsanleitung oder Dokumentation (ReadMe file).

Am Ende dieses Projekts sollte der Entwickler über ein vielseitiges CSV-Reinigungswerkzeug in Form der **CleanerCSV**-Klasse verfügen, das in der Lage ist, eine Vielzahl von Datenreinigungsbedürfnissen einfach und effizient zu erfüllen.

Summary (EN) : DataProcessing (Dynamically)

try to create functions as possible to cover all possible preprocessing actions

- Dropna

- fillna
- replace
- duplicate
- convert to the correct data types
- rename_columns
- etc.

please work also with:

- function-wise
- oop-wise