**Maßnahmen zur Datenbereinigung**

Mithilfe des Microsoft-Excel Programms konnte der Datensatz automatisiert bereinigt werden, da man die Daten beim Einlesen durch bedingte Formatierungen farblich markieren kann, um eventuell kritische Eingabefehler zu überprüfen.

Beispielsweise wurde hier schematisch die Zeitvariable „Jahr“ mit Regeln zum Hervorheben von Zellen definiert, dass jeder Wert der sich nicht innerhalb des Intervalls [1950,2022] befindet rot markiert wird zur gesonderten Überprüfung und gegebenenfalls Bereinigung oder Korrektur.

Bei der dazu abhängigen Variable „Eheschließungen“ wurde die Formatierung so definiert, dass Werte die größer als der Maximalwert sind oder Null ergeben auch gesondert werden.

Gegebenenfalls hier kann man optional definieren, dass Werte die einen gewissen Prozentsatz vom Vorjahr unter – oder übertreten auch farblich markiert werden.

Auf Nachfrage wurde der Datensatz nur bereinigt, sprich: Die fehlerhaften Zeilen gelöscht, und nicht mithilfe der RAW Dateien korrigiert.