1)



Abbildung : Rohdaten

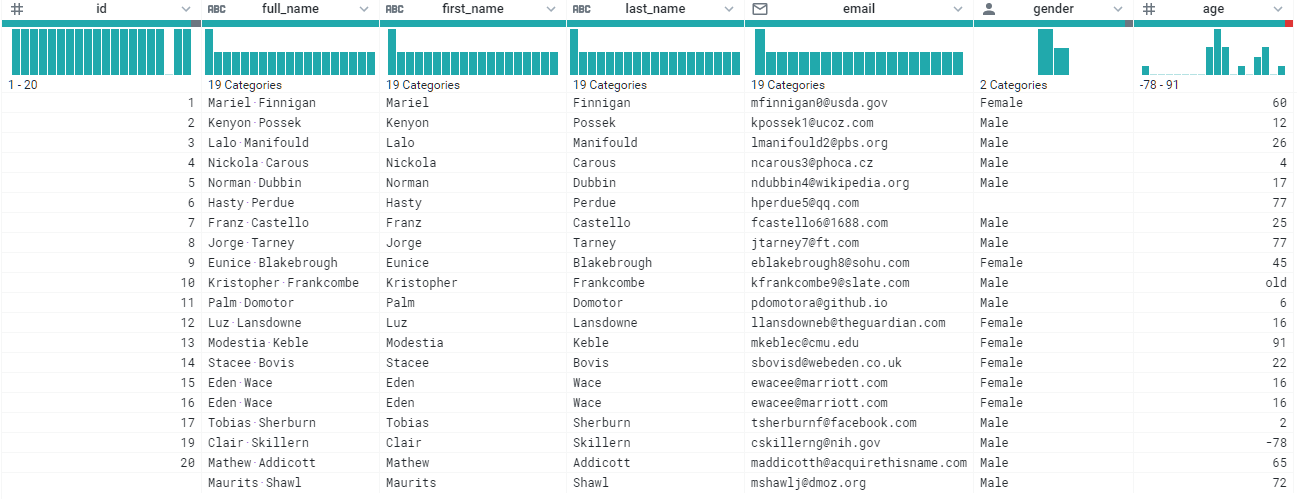


Abbildung : filter type: missing missing: email action: Delete

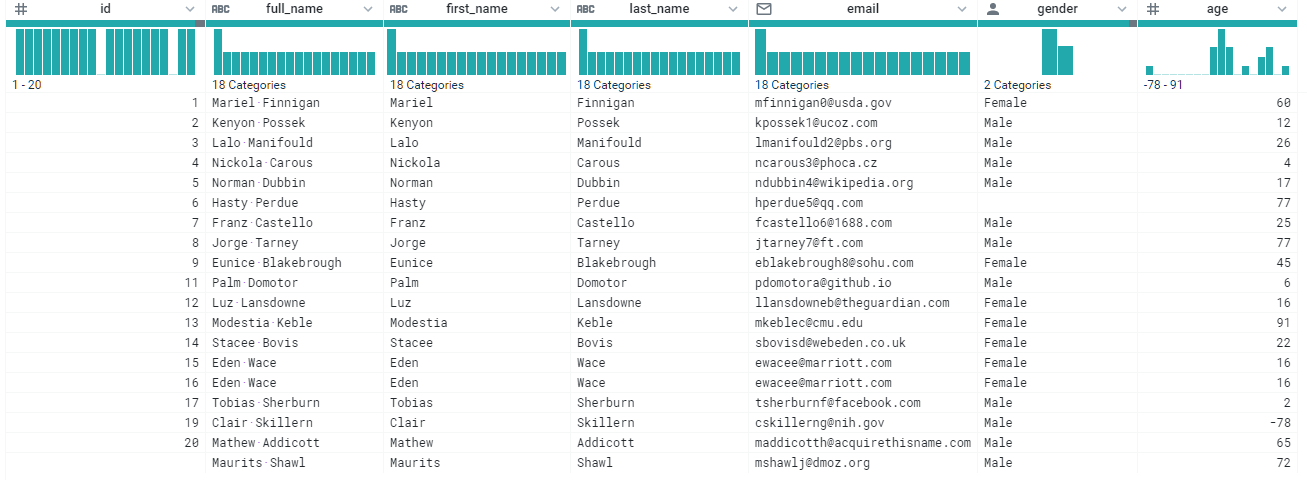


Abbildung : filter type: mismatched col: age mismatched: Integer action: Delete

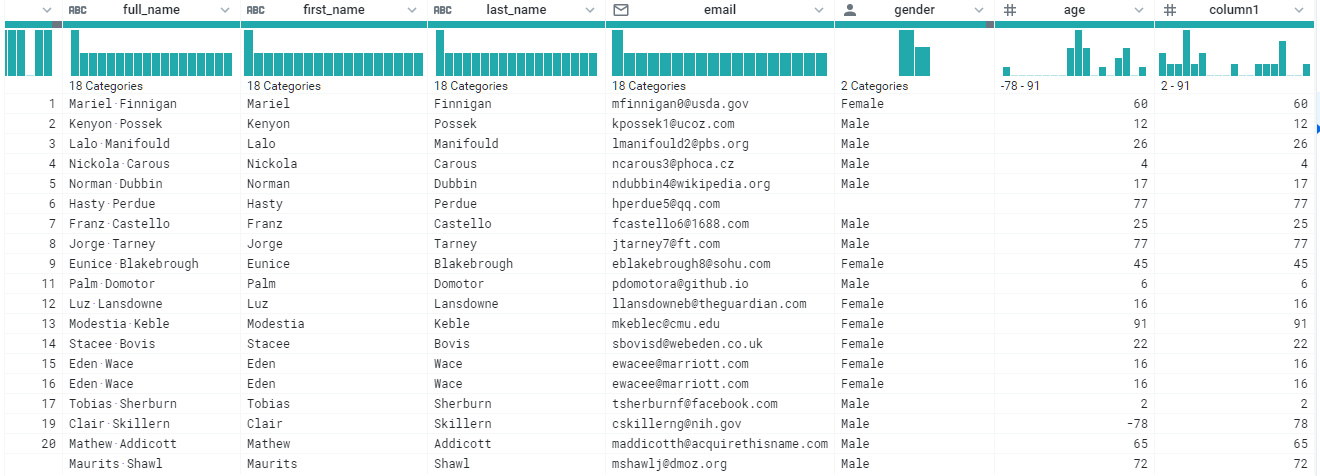


Abbildung : derive type: single value: if(age < 0, age \* -1, age) as: 'column1'

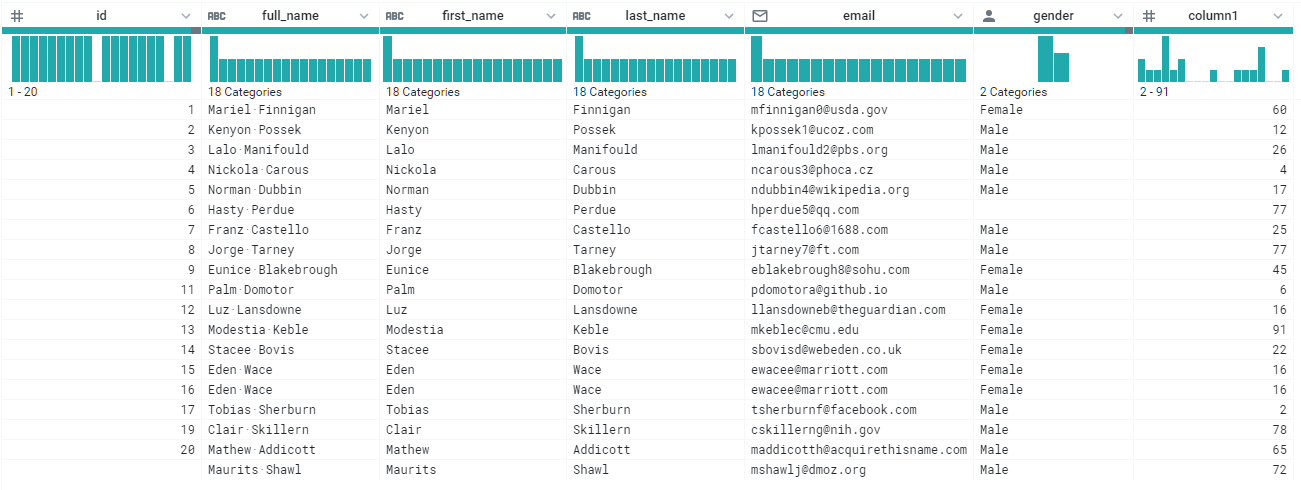


Abbildung : drop col: age action: Drop

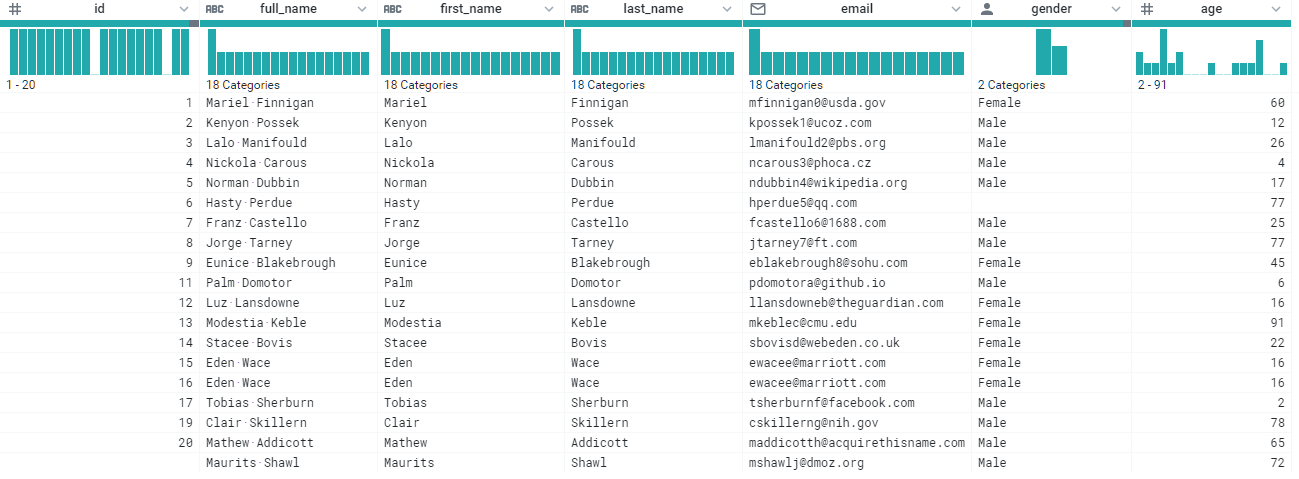


Abbildung : rename type: manual mapping: [column1,'age']

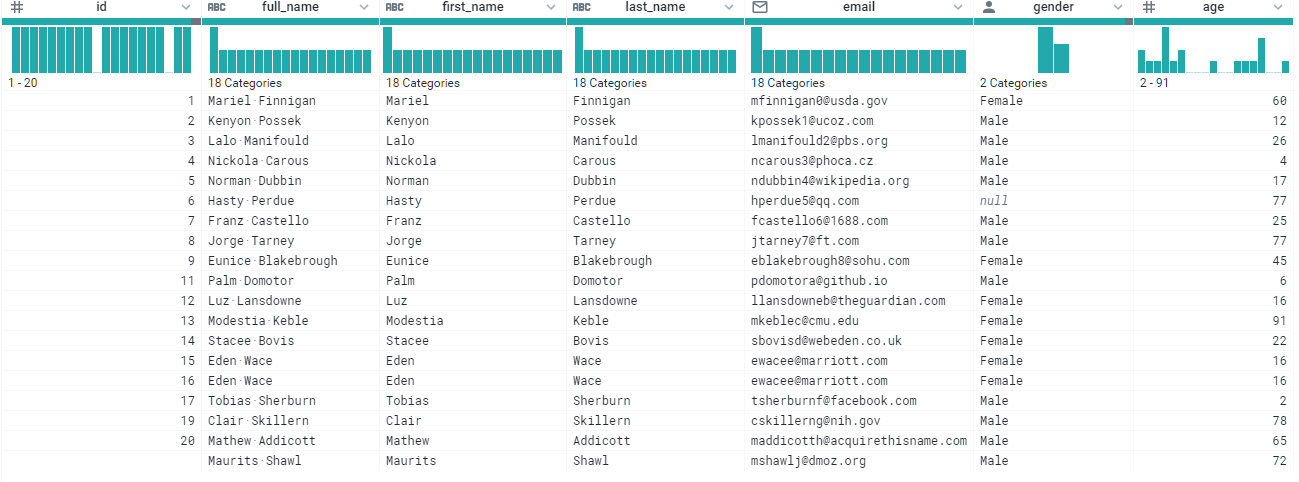


Abbildung : set col: gender value: if(gender == '', null(), $col)

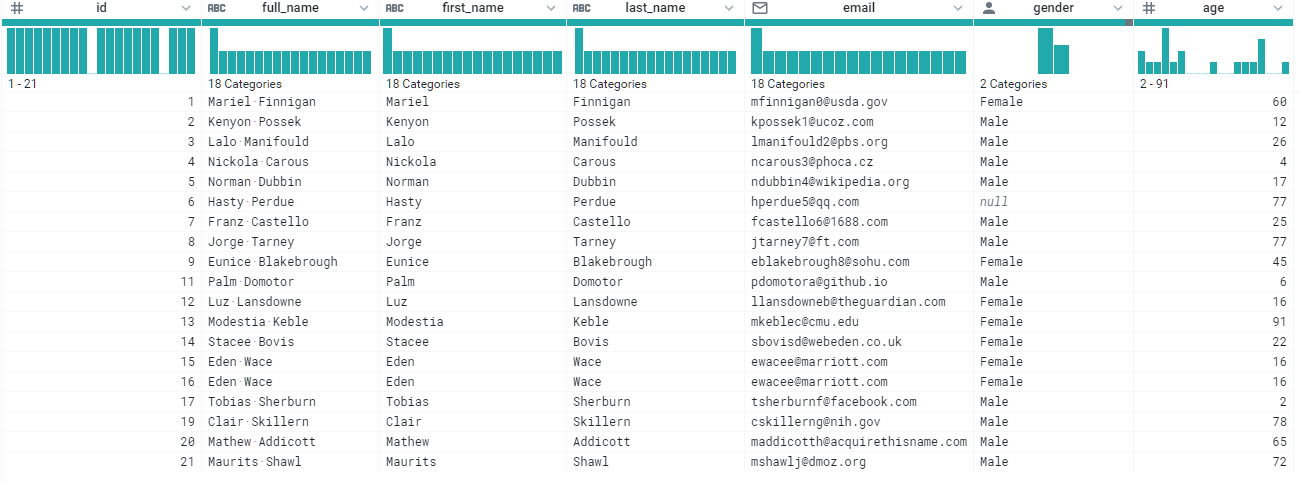


Abbildung : set col: id value: ifmissing($col, max(id) + 1)

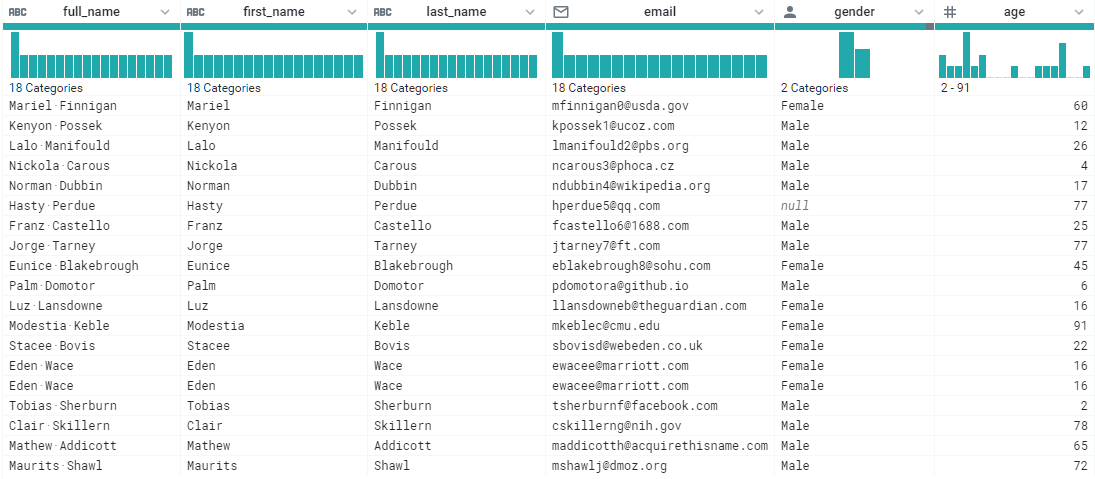


Abbildung : drop col: id action: Drop

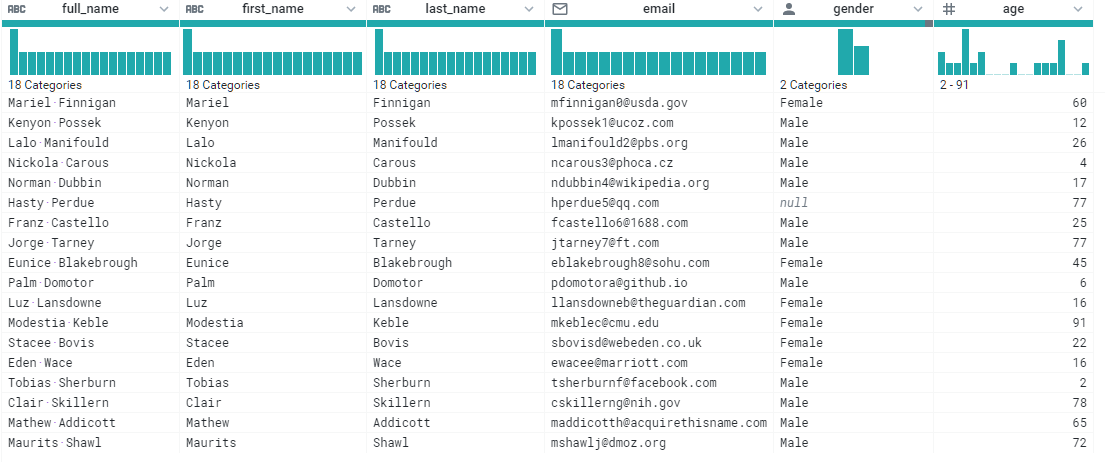


Abbildung : deduplicate

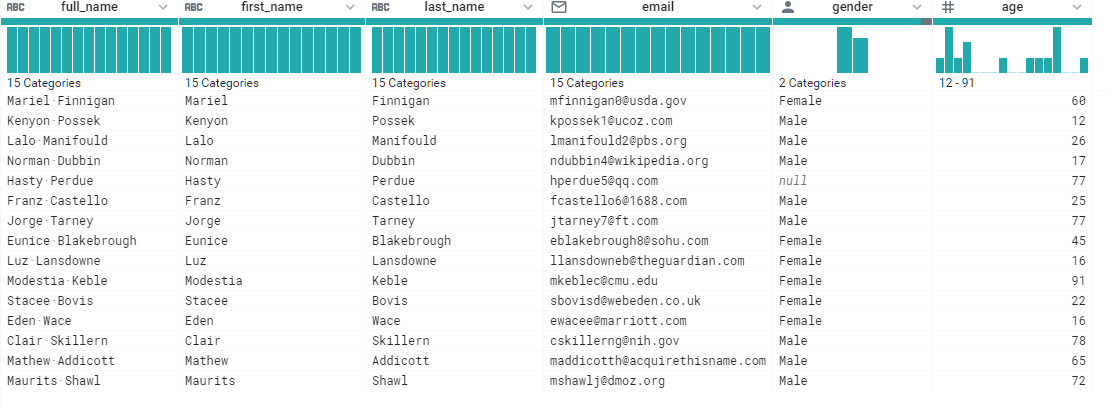


Abbildung : filter type: lessThan col: age lessThan: 10 action: Delete

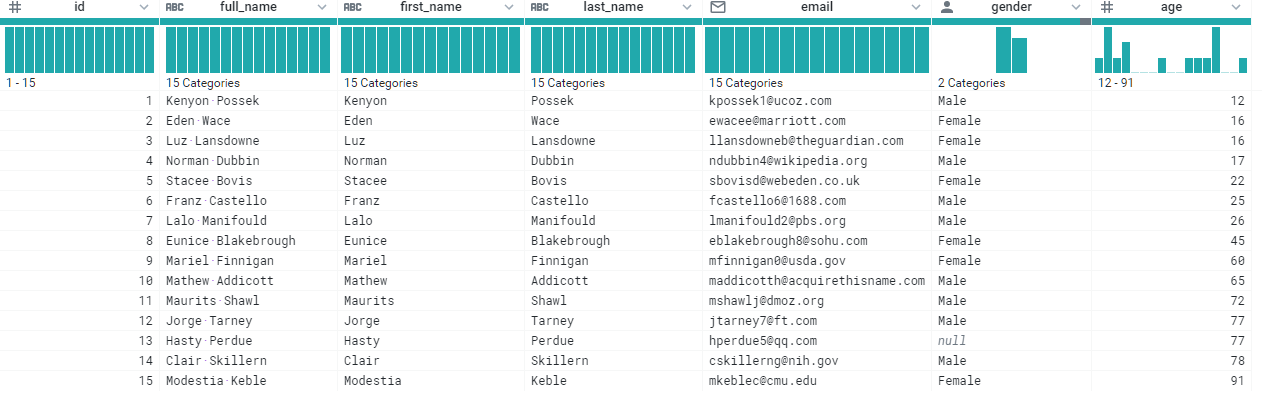


Abbildung : derive type: multiple value: rownumber() order: age as: 'id'

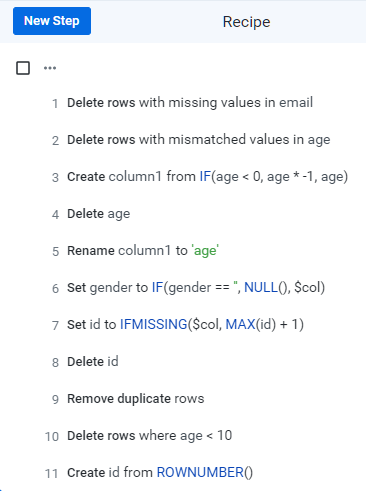


Abbildung : Recipe für Aufgabe 1

2)

1. Schwerwiegende Fehler sind in den Spalten, die Zeit oder Datum enthalten. Diese Zeilen würde ich löschen, da es sich dabei nur um wenige Zeilen (67) handelt.
2. Dann lösche ich in allen Text-Spalten die Anführungszeichen.
3. Aus der Spalte „Number of Customers Affected“ entferne ich alle Kommata, wandle den Datentyp in INT um und ersetze alle Missmatches mit NULL
4. Aus der Spalte „Demand Loss (MW)“ entferne ich alle Buchstaben und Sonderzeichen und wandle sie in INT um, nachdem ich alle MISSMATCHED und MISSING in NULL umgewandelt habe.

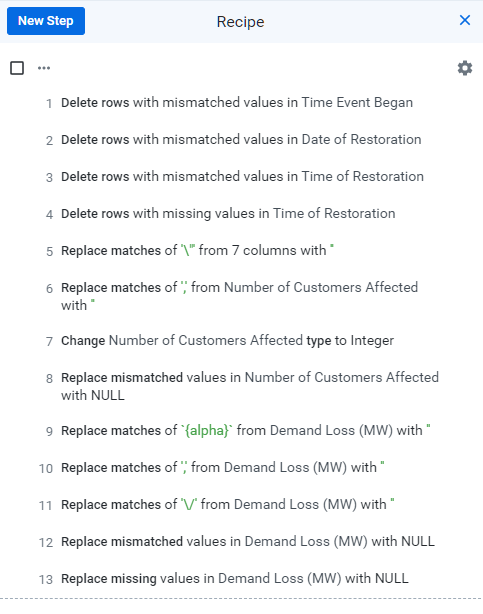


Abbildung 14: Recipe für Aufgabe 2