Data Quality

Haldun Köktaş

2022-11-6

- There are 55 empty cells only in the "CustomerNo" column. 54 of these are related to cancelled orders.

- The data is biased towards "United Kingdom" as the country variable. 485000 of 50000 of purchases made in United Kingdom.

- There are 9994 duplicate rows.

- There are 2 transactions which have a quantity of 80995 and 74215. Both these orders are cancelled. Considering their difference from the rest of the data and they're cancelled, these quantities might not indicate the actual quantity.