Handling missing values

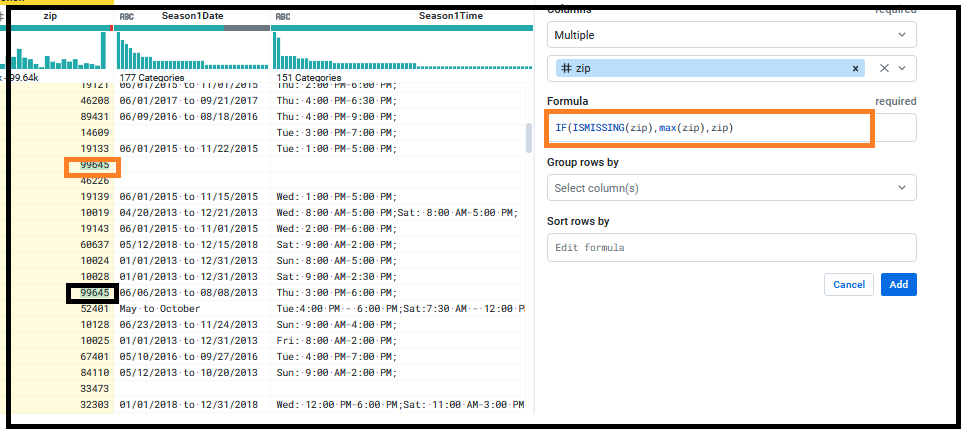
**Handling Missing values**

**Dataset** : Farmers Markets in the United States

<https://www.kaggle.com/datasets/madeleineferguson/farmers-markets-in-the-united-states?resource=download>

Replace missing value with most repeated value.

In dataset ‘zip’ column has a missing value. We can replace null value with max( ).



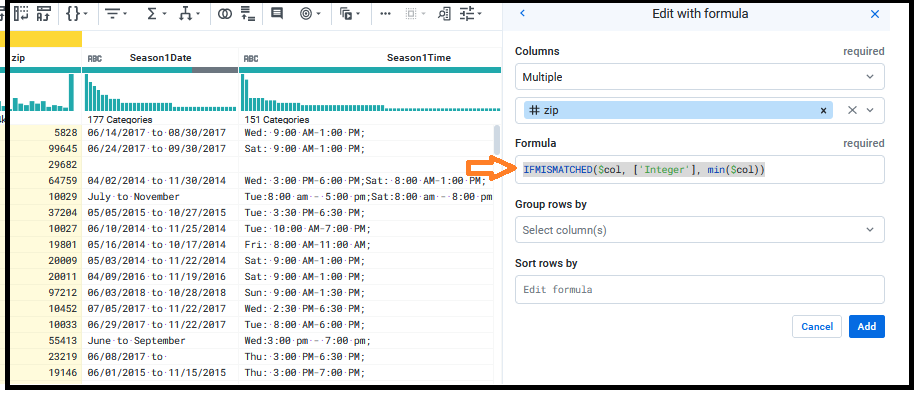
**Remove mismatched value from ‘zip’ column.**

Steps:

Click on the column and go to suggestion box.

Go to Edit with formula and write the below formula:

IFMISMATCHED($col, ['Integer'], min($col))



Click on Add.

Q. **check the zip code as it should be a 5 digit number.**

Select the 4 digit number from column ‘zip’ .

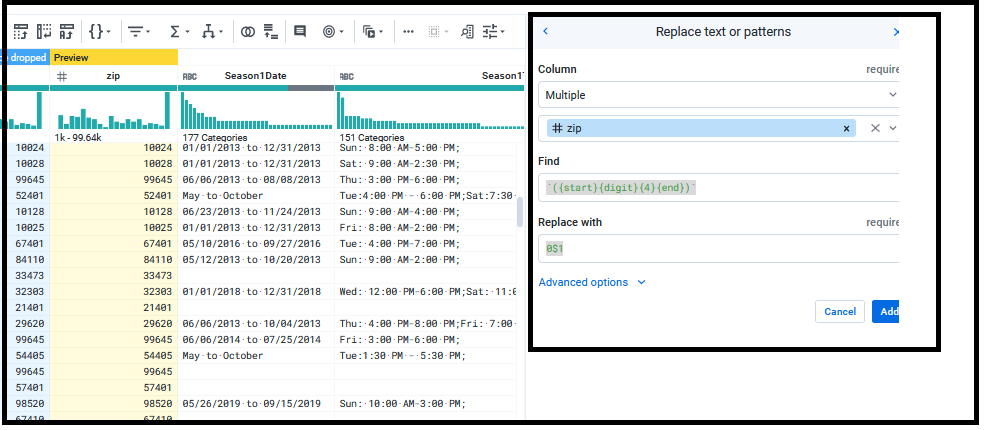
Go to suggestion box

Select Replace option and will get options as

* Column
  + Multiple
* Give Column name
* Find
  + `({start}{digit}{4}{end})`

It explains that the number is a 4 digit number and there is nothing before and after the number.

* Click on Edit and write 0$1 in Replace with
* Click add



**Q**. **Replace empty rows with NA values for ‘facebook ‘ column.**

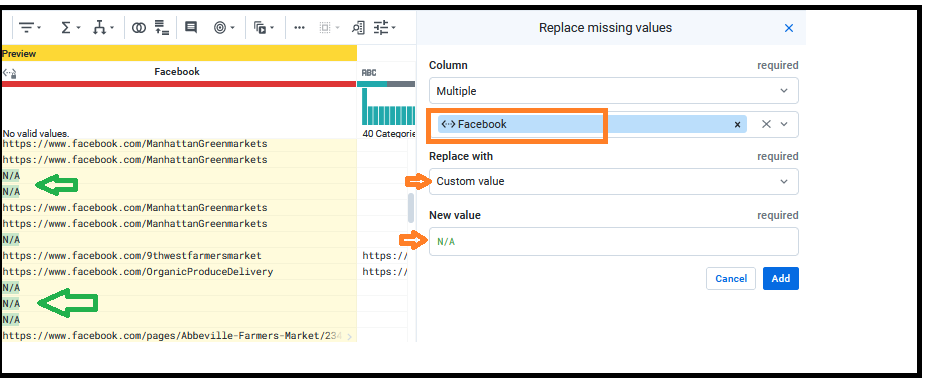
Go to replace option…

Select Missing option

Select Custom value

On RHS of suggestion box we will set a value for New value column.

Click Add



END