Join

Join combine columns based on column key. Join combines multiple datasets with the help of column key.

* Go to Recipe and click on New step
* Search join
* Select datasets which you want to take for join. It has following options
* Inner
  + - * It returns only the matching rows from both the tables.
* Left
  + - * It returns all matching rows and matching rows from left table.
* Outer
  + - * It returns all the rows from both the tables whether it is matching or not.
* Cross
* right
  + - * It returns all matching rows and matching rows from right table.

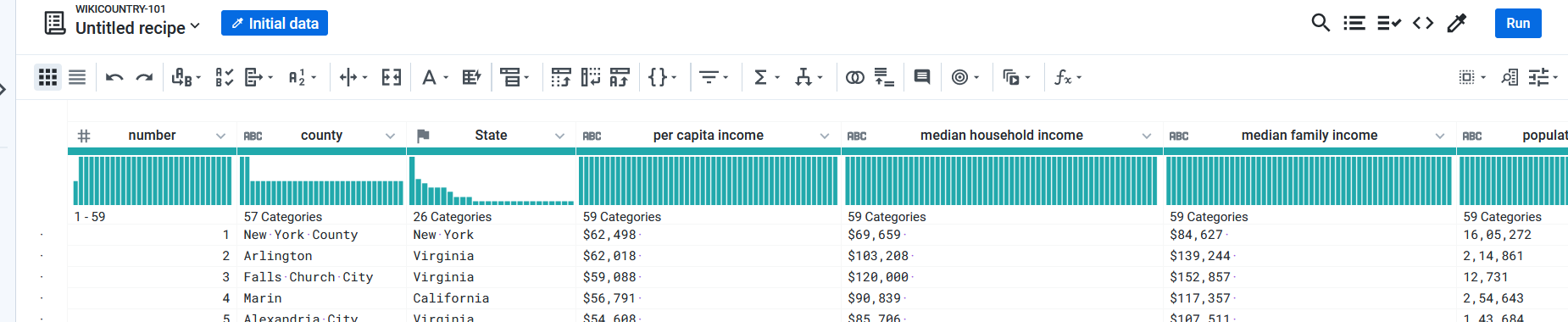


**Dataset**

[**https://www.kaggle.com/code/benbrooker1/u-s-farmer-s-markets-trend-analysis/data?select=farmers\_markets\_from\_usda.csv**](https://www.kaggle.com/code/benbrooker1/u-s-farmer-s-markets-trend-analysis/data?select=farmers_markets_from_usda.csv)

Applying join on farmers\_markets\_from\_usda.csv and wiki\_county\_info.csv datasets.

* Take the column key as **country**.

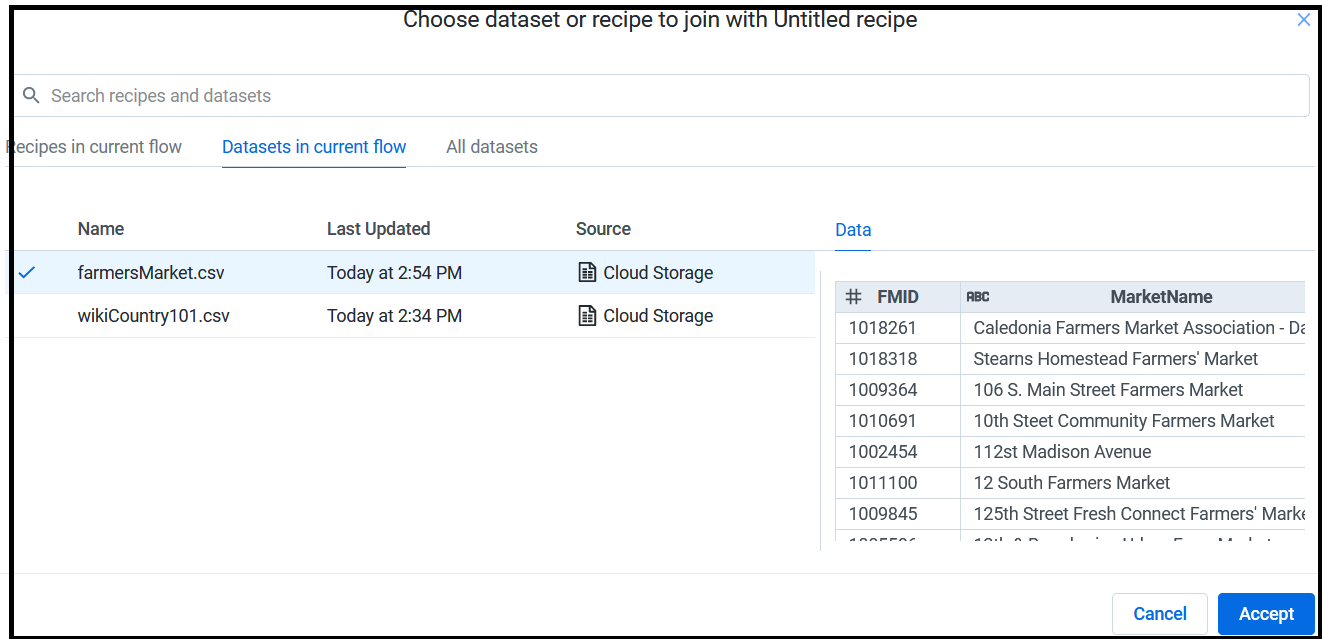


Steps

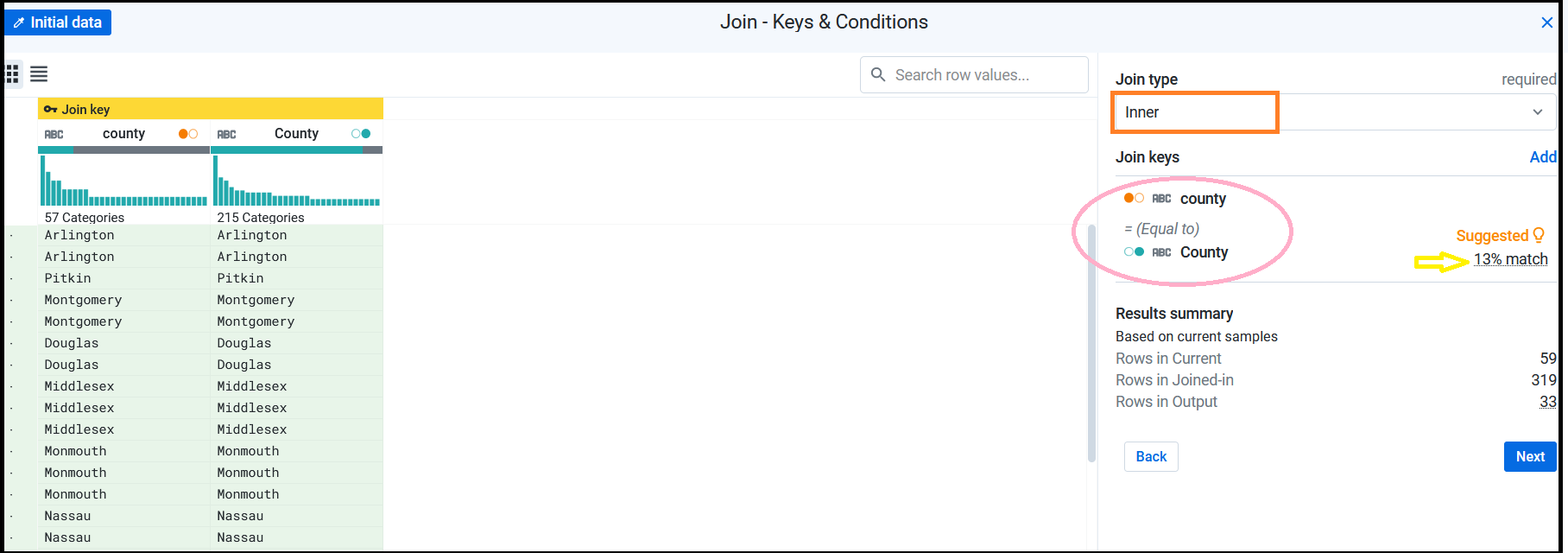
Click on join option.

Select the dataset from dataset in current flow.

**Inner join**



The result shows



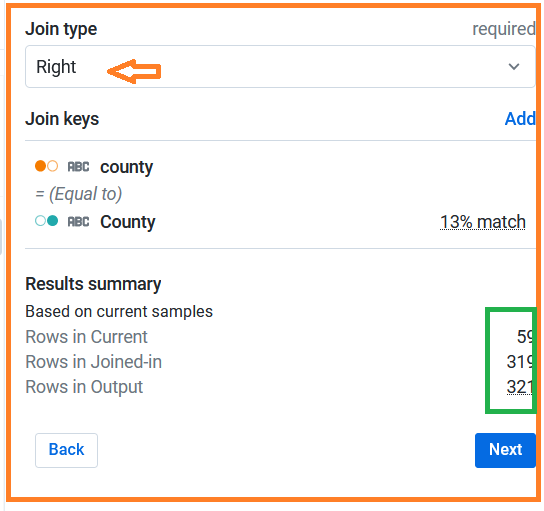
Rows in original dataset was 59

Rows after join becomes 319

13% match and number of matching rows is 33

Click on Next

We will get a set of column names with blue one from left dataset and orange colour one from right dataset.



We can take the option join and check the result.

Here we have

* Rows in current 59
* Rows in joined-in 319
* Rows in Output 321(it returns matching values with records from right table).

end

* Click on Add to Recipe.

**Remove any symbol from column values**

* Go to NEW Recipe
* Search for text format….do select the column and click add.

**Change data types**

* Click on change type option (select the column and from drop down menu).