

Parse Fixed-Width File and Infer Columns

For datasets that have a fixed width for each row, determining the column breaks can be more challenging, due to the uncertain number of spaces and tabs between each data element. With enhanced pattern matching, the application can help you identify the appropriate locations to break columns and then trim down the data to eliminate the whitespace padding.

Steps:

1. Import your fixed-width dataset through the application and begin wrangling.
2. The data should now look similar to the following:

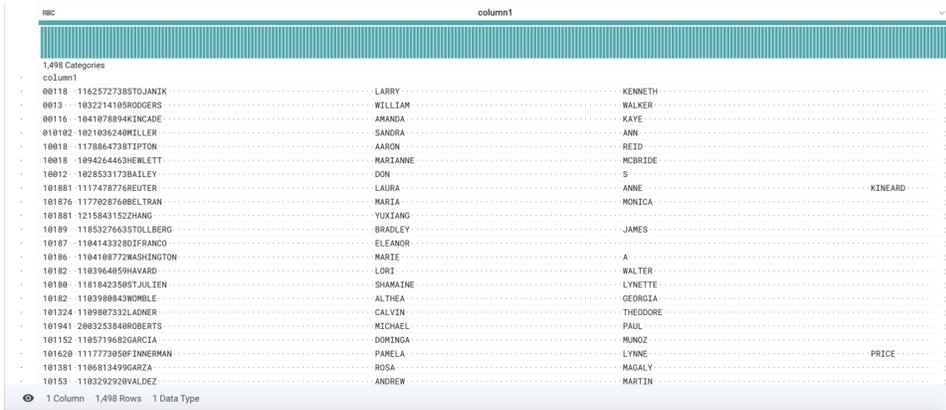


Figure: Fixed-width dataset after import

3. From the drop-down to the right of the column name, select **Column Details**.
4. In the Column Details panel, click the Patterns tab.
5. Click in the All Patterns area.

NOTE: Selecting a specific pattern token will generate suggestions for only that particular token.

NOTE: If the application has inferred that the dataset is fixed-width, then the All Patterns area is the only available selection. If the dataset is not inferred as fixed-width, you should see multiple categories of patterns.

6. From the suggestion cards, click the Split one.
7. Close the Column Details panel.
8. In the Transform preview window, verify that the column splits look ok.
 - a. If a column contains multiple columns of data, click **Edit**.
 - b. Verify that you are splitting based on position numbers, which means that column splits are done based on the number of characters from the left side of each line.
 - c. Your recipe step might look similar to the following:

| | |
|-----------------------------------|---|
| Transformation Name | Split columns by positions |
| Parameter: Column to split | column1 |
| Parameter: Option | By positions |
| Parameter: Positions | 7, 67, 117, 167, 217, 221, 239, 251, 253, 303, 315, 317, 329, 341, 391, 400, 512, 560, 610, 630, 650, 660 |

- d. In the list of values for positions, insert a new position number for the column or columns that contain multiple columns of data.
 - e. Verify your changes in the Transform Preview panel.
9. Click **Add**.
 10. Verify that the columns are split correctly.
 11. You can use the following step to remove the whitespace from each cell value.

| | |
|----------------------------|--------------------------|
| Transformation Name | Edit column with formula |
| Parameter: Column | * |
| Parameter: Formula | TRIM(\$col) |

12. Click **Add**.