

# Create Class

Create class attribute from a string attribute.

## Inputs

- Data: input dataset

## Outputs

- Data: dataset with a new class variable

**Create Class** creates a new class attribute from an existing discrete or string attribute. The widget matches the string value of the selected attribute and constructs a new user-defined value for matching instances.

1

From column: iris

2

Name	Substring	#Instances
× C1	(remaining instan...	150
× C2	(unused)	
+		

3

Name for the new class:

4

☐ Match only at the beginning  
☐ Case sensitive

5

Report

6

Apply

1. The attribute the new class is constructed from.

## 2. Matching:

- Name: the name of the new class value
- Substring: regex-defined substring that will match the values from the above-defined attribute
- Instances: the number of instances matching the substring
- Press '+' to add a new class value

## 3. Name of the new class column.

4. Match only at the beginning will begin matching from the beginning of the string. Case sensitive will match by case, too.

5. Produce a report.

6. Press *Apply* to commit the results.

## Example

Here is a simple example with the *auto-mpg* dataset. Pass the data to **Create Class**. Select *car\_name* as a column to create the new class from. Here, we wish to create new values that match the car brand. First, we type *ford* as the new value for the matching strings. Then we define the substring that will match the data instances. This means that all instances containing *ford* in their *car\_name*, will now have a value *ford* in the new class column. Next, we define the same for *honda* and *fiat*. The widget will tell us how many instance are yet unmatched (remaining instances). We will name them *other*, but you can continue creating new values by adding a condition with '+'.  
We named our new class column *car\_brand* and we matched at the beginning of the string.

**Create Class**

From column: **car\_name**

Name	Substring	#Instances
× ford	ford	51
× honda	honda	13
× fiat	fiat	8
× other	(remaining instan...	326 + 72

+

Name for the new class: **car\_brand**

☒ Match only at the beginning  
☐ Case sensitive

**Report** **Apply**

**Data Table**

Info

398 instances  
 8 features (0.2% missing values)  
 Discrete class with 4 values (no missing values)  
 1 meta attribute (no missing values)

Variables

☒ Show variable labels (if present)  
☐ Visualize continuous values  
☒ Color by instance classes

Selection

☒ Select full rows

**Restore Original Order**

**Report**

☒ **Send Automatically**

	car_brand	mpg	cylinders	displacement	horsepower
147	other	28.000	4	90.000	75.000
148	fiat	24.000	4	90.000	75.000
149	fiat	26.000	4	116.000	75.000
150	honda	24.000	4	120.000	97.000
151	other	26.000	4	108.000	93.000
152	fiat	31.000	4	79.000	67.000
153	other	19.000	6	225.000	95.000
154	other	18.000	6	250.000	105.000
155	other	15.000	6	250.000	72.000
156	ford	15.000	6	250.000	72.000
157	other	16.000	8	400.000	170.000
158	other	15.000	8	350.000	145.000
159	other	16.000	8	318.000	150.000
160	ford	14.000	8	351.000	148.000
161	other	17.000	6	231.000	110.000
162	other	16.000	6	250.000	105.000
163	other	15.000	6	258.000	110.000
164	other	18.000	6	225.000	95.000
165	other	21.000	6	231.000	110.000

Finally, we can observe the new column in a **Data Table** or use the value as color in the **Scatter Plot**.