Parsing an HTML Web Page as Text

Content Created by EXCELGURU
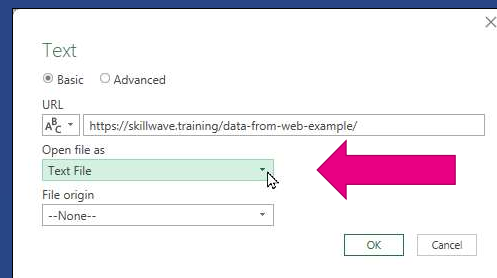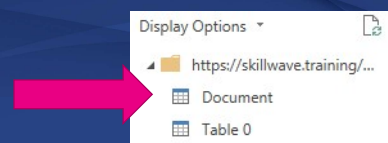
© Skillwave Training

If your Navigator window shows "Document"

Select Document -> Transform Data
Delete the Navigation step
Click the Gear icon beside Source
Change "Open file as" to Text File

Display Options ▼

▲ 📁 https://skillwave.training/...
   ▦ Document
   ▦ Table 0

Text

◉ Basic ○ Advanced

URL
A^BC ▼ | https://skillwave.training/data-from-web-example/

Open file as
Text File ▼

File origin
--None--

OK   Cancel

© Skillwave Training

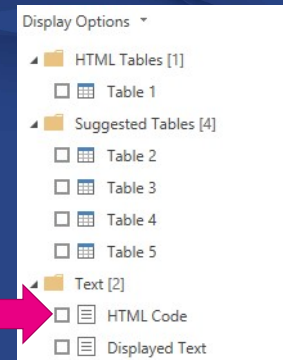# If your Navigator window shows HTML Code

Select HTML Code -> Transform Data

Click the *fx* beside the formula bar

Replace the formula with:
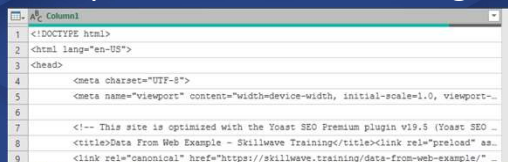
```
Lines.FromText(#"HTML Code")
```

Convert the output to a Table (using the default options)

Parsing a Web Page (as Text)

Your output should now be a single column Table:

The two methods shown here result in different interpretations of line breaks. This results in differing line numbers and steps used to clean the data!

Helpful Tips

Add an Index Column (from 0)

Filter to identify a term where your data starts (note the Index number)

Remove the filter, then remove Top x Rows (using the number from above)

Replace Values and Split Column by Delimiter are your best friends!