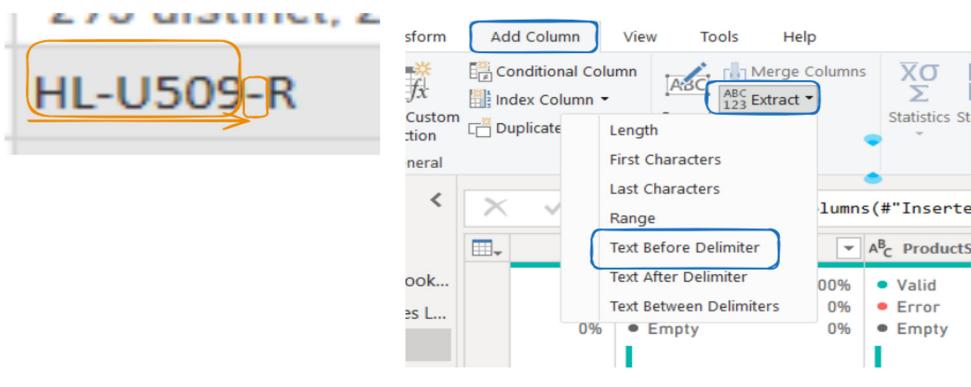


## Power BI - Data Transformation - p3

### Assignment : Transforming Data

1. Create queries to connect to the two new .csv files
2. Name your queries Product Category Lookup and Product Subcategory Lookup
3. Confirm that column headers have been promoted and that all data types are correct
4. Add a new column to extract all characters before the dash ("") in the Product SKU column, and name it "SKU Type"
5. Update the SKU Type calculation above to return all characters before second dash, instead of the first
6. Replace zeros (0) in the Product Style column with "NA"
7. Close and load to your data model



#### Text Before Delimiter

Enter the delimiter that marks the end of what you would like to extract.

Delimiter

Advanced options

Scan for the delimiter

Number of delimiters to skip

OK

Cancel

6.

**ProductStyle**

- Valid 100%
- Error 0%
- Empty 0%

4 distinct, 0 unique

Replace Values

Replace one value with another in the selected columns.

Value To Find: 0

Replace With: NA

Advanced options

Search

(Select All)

M

NA

U

W

## EXTRACTING DATA FROM THE WEB

- Power Query includes a native Web connector for importing web-hosted files (csv, xlsx, etc.) or scraping URLs for anything that Power Query can identify as a structured table.

## DATA PROFILING: COLUMN QUALITY

Profiling tools like column quality, column distribution, and column profile allow you to explore the quality, composition, and distribution of your data before loading it into the Power BI front-end.

**Prefix**

- Valid 99%
- Error 0%
- Empty < 1%

Mr.  
Mr.

**FirstName**

Category	Percentage	Count
Valid	99%	991
Error	0%	0
Empty	< 1%	9

Remove Empty

...

Copy  
Keep Duplicates  
Keep Errors  
Remove Duplicates  
Remove Empty  
Remove Errors  
Replace Errors

Hover over the column quality box to see the number of records in each category

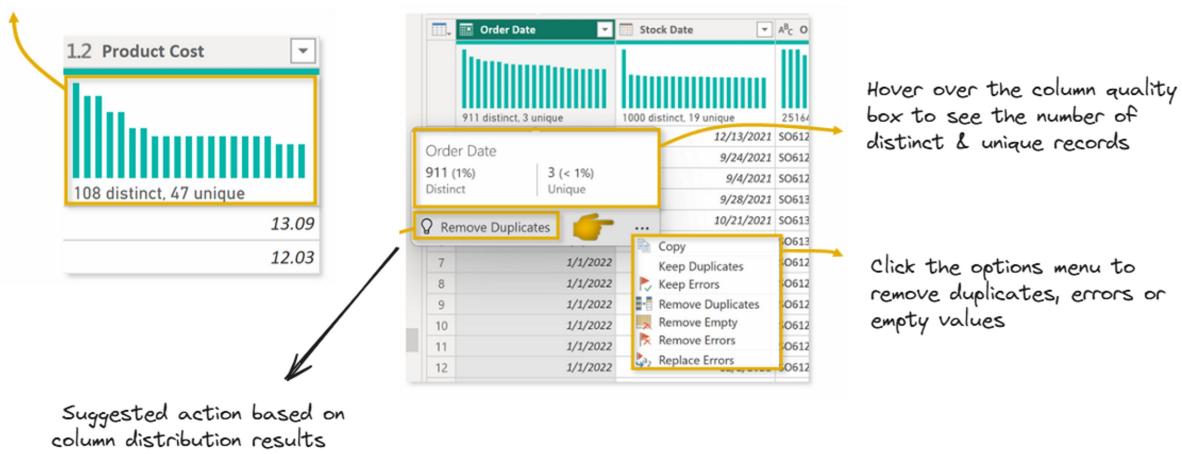
Column quality shows the percentage of values within a column that are valid, contain errors, or are empty

Click the options menu to remove duplicates, errors or empty values

Profiling tools are a great way to quickly find and address common data quality issues in one place, instead of having to manually apply multiple tools or filters

## COLUMN DISTRIBUTION

Column distribution provides a sample distribution of the data in a column



## COLUMN PROFILE

**Column profile** provides a more holistic view of the data in a column, including a sample distribution and profiling statistics

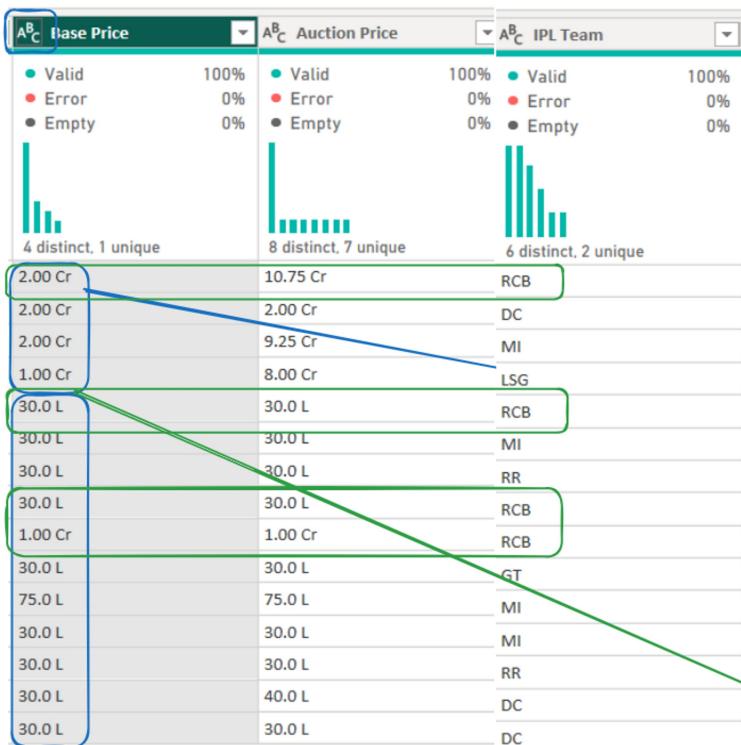
**Column statistics** provide more detailed profiling metrics, including:

- Count = 293**  
(total number of values in column)
- Distinct Count = 119**  
(total number of distinct values, whether they appear once or multiple times)
- Unique = 82**  
(total number of values that appear exactly once)
- Min & Max**  
(lowest and highest observed values)

**Note:** Typically only useful for numerical values

**Value distribution**

**Hover over the value distribution bar for suggested transformations and additional options**



3Cr  
60 lakh

RCB

Manipulation of result. Aggregation.

$$30+30+2+1 = 33$$

$$\begin{array}{r} 20000000 \\ + 3000000 \\ \hline \end{array}$$

Player Name	Player Type	Country	Base Price	Auction Price	IPL Team
Bhuvneshwar Kumar	Bowler	India	2000000	10750000	RCB
Faf du Plessis	Batter	South Africa	2000000	2000000	DC
Deepak Chahar	Bowler	India	2000000	9250000	MI
Akash Deep	Bowler	India	1000000	8000000	LSG
Mohit Rathee	Bowler	India	3000000	3000000	RCB
Vignesh Puthur	Allrounder	India	3000000	3000000	MI
Ashok Sharma	Bowler	India	3000000	3000000	RR
Abhinandan Singh	Bowler	India	3000000	3000000	RCB
Lungi Ngidi	Bowler	South Africa	1000000	1000000	RCB
Kulwant Khejroliya	Bowler	India	3000000	3000000	GT
Lizaad Williams	Bowler	South Africa	7500000	7500000	MI
Arjun Tendulkar	Bowler	India	3000000	3000000	MI
Kunal Singh Rathore	WK-Batter	India	3000000	3000000	RR
Madhav Tiwari	Allrounder	India	3000000	4000000	DC
Tripurana Vijay	Allrounder	India	3000000	3000000	DC

IPL Team	Count of Auction Price
DC	3
GT	1
LSG	1
MI	4
RCB	4
RR	2
<b>Total</b>	<b>15</b>

IPL Team	Sum of Auction Price
DC	27000000
GT	3000000
LSG	80000000
MI	106000000
RCB	123500000
RR	6000000
<b>Total</b>	<b>345500000</b>

Aggregated Columns

Group BY - Team