Load, Clean, and Transform Data in Power BI - I: Takeaways

Syntax

- Promoting headers (method one):

  ```
  1. Select the row

  2. From the Home tab, click on  Use First Row as Headers
  ```

- Promoting headers (method two):

  ```
  1. Select the row

  2. Select the drop-down button next to the column

  3. Select Use First Row as Headers.
  ```

- Renaming columns (method one):

  ```
  1. Right click the header

  2. Select rename

  3. Edit the name and press Enter
  ```

- Renaming columns (method two):

  ```
  1. Double click the column header

  2. Overwrite the existing name
  ```

- Removing top rows:

  ```
  1. Select the rows to remove

  2. On the Home tab, Click on Remove Rows

  3. Select Remove Top Rows
  ```

- Removing columns (method one):

  ```
  1. Select the columns to remove

  2. On the Home tab, Click on Remove Columns

  3. Select Remove Columns
  ```

- Removing columns (method two):

  ```
  1. Select the columns to keep

  2. On the Home tab, Click on Remove Columns

  3. Select Remove Other Columns
  ```

- Pivoting columns:

  ```
  1. Go to the Transform tab

  2. Click on Pivot Column

  3. Select the column from the Values Column field

  4. Expand the Advanced options

  5. Select the aggregate function from the Aggregate Value Function field

  6. Select Close & Apply
  ```

- Unpivoting columns:

```
1. Select the columns to unpivot

2. Go to the Transform tab

3. Click on Unpivot Columns

4. Rename (if needed) the new Attribute and Value columns
```

Concepts

- Advantages of cleaning data:

  - Measures and columns produce more accurate results when they perform aggregations and calculations.

  - Tables are organized, where users can find the data in an intuitive manner.

  - Duplicates are removed, making data navigation simpler. It will also produce columns that can be used in slicers and filters.

  - A complicated column can be split into two, simpler columns. Multiple columns can be combined into one column for readability.

  - Codes and integers can be replaced with human readable values.

- Shaping the initial data means:

  - Renaming columns or tables

  - Changing text to numbers

  - Removing rows

  - Setting the first row as headers

- Pivoting the data consists of summarizing it using different math functions: count, minimum, maximum, median, average, sum.

References

- Data types in Power BI Desktop

- Auto generated data type documentation

- Shape and Combine Data in Power BI documentation