**abstract:** This column contains a brief summary of the article. It gives a quick overview of the content.

**web\_url**: This is the URL to the actual article.

**lead\_paragraph:** The lead paragraph of the article, which usually introduces the main focus or topic of the piece.

**multimedia:** This column likely contains multimedia-related data, such as images or videos associated with the article.

**headline:** Information about the article's headline

**keywords:** A list of keywords or tags associated with the article. These might be used for categorization or search optimization.

**pub\_date:** The publication date and time of the article.

**document\_type**: This column indicates the type of document. For example, it might distinguish between standard articles, blog posts, etc.

**news\_desk:** This refers to the specific department or section of the news organization that produced the article, like 'National', 'Culture', 'Foreign', etc.

**section\_name:** The section of the newspaper in which the article appeared, such as 'U.S.', 'Arts', 'World', 'Health', etc.

**byline**: Information about the author(s) of the article. This could include the author's name and possibly other details like their role or contributions.

**type\_of\_material:** This column describes the nature of the content, like 'News', 'Obituary (Obit)', 'Op-Ed', etc.

**\_id:** This appears to be a unique identifier for each article, possibly used internally by the New York Times for tracking or database management.

**word\_count:** The number of words in the article.

**uri:** A URI (Uniform Resource Identifier) that uniquely identifies the article. It might be used internally for referencing articles in the NYT system.