popular databases that are used widely for text data. Each database comes with ways to query and perform operations on text. We'll implement some of these operations. Finally, we will introduce the concept of data maintenance and discuss some useful tips and tricks to prevent the data from corruption.

This chapter includes the following sections.

- Sources of data
- Data extraction
- Data storage

1.2 SOURCES OF DATA

1.2.1 Generated by businesses

The most common source of text data is the data generated by the business's operations and is dependent on what the business does. For example, in real estate, sources of text data include property listing descriptions, agent comments, legal documents, and customer interaction data. For some other industry verticals, the source of text data can include social media posts, product descriptions, articles, web documents, chat data, or a combination thereof. When there is an absence of owned (first-party) data, organizations leverage data from different vendors and clients. Overall, from a business's standpoint, the commonly seen text data is of the following types.

1. Customer reviews/comments

   User comments are a very common source of text, especially from social media, e-commerce, and hospitality businesses that collect product reviews. For instance, Google and Yelp collect reviews across brands as well as small and large businesses.

2. Social media/blog posts

   Social media posts and blogs find presence in most types of businesses. In today's world, social media reaches more people globally than any other form of media. Whether or not a business is directly associated with social media, there's often social media presence of businesses, products, service promotions, articles, or more. On the other hand, there are many businesses offering a product/service for analyzing the social media presence of other businesses. Whether one is gathering one's own media data or on behalf of a client, there's a rich volume of text associated which makes this a popular text data source that spans many industry verticals.

3. Chat data

   E-commerce, banking, and many other industries leverage chat data. Chat data is essentially the chat history of messages exchanged between a business and its client or customer. This is a common source of text data in industries and is