Appendix A

This appendix explains how to prepare the sample data so that you can practice the exercises in this book.

The process is as follows:

1. Download the sample data from the GitHub repository and save the content to a local folder.
2. Run the Databricks interface (either the commercial version or the Community version)
3. Upload all the sample .csv files. Do this by clicking
   1. Data (in the left menu)
   2. Create Table
   3. Browse to from the folder that you created in (1) above and select the .CSV files only
4. Create a new notebook
5. Click File=> Import Notebook and browse to the notebook file LoadData.jpynb in the folder that you created in (1) above.
6. Click Run => Run All to run the notebook and prepare the sample data.