Learn data science by doing data science!

*Data Science Using Python and R* will get you plugged into the world's two most widespread open-source platforms for data science: Python and R.

Data science is hot. *Bloomberg* called data scientist “the hottest job in America.” Python and R are the top two open-source data science tools in the world. In *Data Science Using Python and R*, you will learn step-by-step how to produce hands-on solutions to real-world business problems, using state-of-the-art techniques.

*Data Science Using Python and R* is written for the general reader with no previous analytics or programming experience. An entire chapter is dedicated to learning the basics of Python and R. Then, each chapter presents step-by-step instructions and walkthroughs for solving data science problems using Python and R.

Those with analytics experience will appreciate having a one-stop shop for learning how to do data science using Python and R. Topics covered include data preparation, exploratory data analysis, preparing to model the data, decision trees, model evaluation, misclassification costs, naïve Bayes classification, neural networks, clustering, regression modeling, dimension reduction, and association rules mining.
Further, exciting new topics such as random forests and general linear models are also included. The book emphasizes data-driven error costs to enhance profitability, which avoids the common pitfalls that may cost a company millions of dollars.

*Data Science Using Python and R* provides exercises at the end of every chapter, totaling over 500 exercises in the book. Readers will therefore have plenty of opportunity to test their newfound data science skills and expertise. In the Hands-on Analysis exercises, readers are challenged to solve interesting business problems using real-world data sets.

---

### ABOUT THE AUTHOR

**CHANTAL D. LAROSE, P HD,** is an Assistant Professor of Statistics & Data Science at Eastern Connecticut State University (ECSU). She has co-authored three books on data science and predictive analytics and helped develop data science programs at ECSU and SUNY New Paltz. Her PhD dissertation, *Model-Based Clustering of Incomplete Data*, tackles the persistent problem of trying to do data science with incomplete data.

**DANIEL T. LAROSE, P HD,** is a Professor of Data Science and Statistics and Director of the Data Science programs at Central Connecticut State University. He has published many books on data science, data mining, predictive analytics, and statistics. His consulting clients include *The Economist* magazine, *Forbes* Magazine, the CIT Group, and Microsoft.

---

### RELATED RESOURCES

<table>
<thead>
<tr>
<th>Student</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Student Companion Site</td>
<td>View Instructor Companion Site</td>
</tr>
</tbody>
</table>

---

### SERIES

**Wiley Series on Methods and Applications in Data Mining**

To purchase this product, please visit [https://www.wiley.com/en-us/9781119526810](https://www.wiley.com/en-us/9781119526810)