
Preface

Subject Area of the Book

In this era, where a huge amount of information from different fields is gathered and stored, its analysis and the extraction of value have become one of the most attractive tasks for companies and society in general. The design of solutions for the new questions emerged from data has required multidisciplinary teams. Computer scientists, statisticians, mathematicians, biologists, journalists and sociologists, as well as many others are now working together in order to provide knowledge from data. This new interdisciplinary field is called *data science*.

The pipeline of any data science goes through asking the right questions, gathering data, cleaning data, generating hypothesis, making inferences, visualizing data, assessing solutions, etc.

Organization and Feature of the Book

This book is an introduction to concepts, techniques, and applications in data science. This book focuses on the analysis of data, covering concepts from statistics to machine learning, techniques for graph analysis and parallel programming, and applications such as recommender systems or sentiment analysis.

All chapters introduce new concepts that are illustrated by practical cases using real data. Public databases such as Eurostat, different social networks, and MovieLens are used. Specific questions about the data are posed in each chapter. The solutions to these questions are implemented using Python programming language and presented in code boxes properly commented. This allows the reader to learn data science by solving problems which can generalize to other problems.

This book is not intended to cover the whole set of data science methods neither to provide a complete collection of references. Currently, data science is an increasing and emerging field, so readers are encouraged to look for specific methods and references using keywords in the net.

Target Audiences

This book is addressed to upper-tier undergraduate and beginning graduate students from technical disciplines. Moreover, this book is also addressed to professional audiences following continuous education short courses and to researchers from diverse areas following self-study courses.

Basic skills in computer science, mathematics, and statistics are required. Code programming in Python is of benefit. However, even if the reader is new to Python, this should not be a problem, since acquiring the Python basics is manageable in a short period of time.

Previous Uses of the Materials

Parts of the presented materials have been used in the postgraduate course of *Data Science and Big Data* from Universitat de Barcelona. All contributing authors are involved in this course.

Suggested Uses of the Book

This book can be used in any introductory data science course. The problem-based approach adopted to introduce new concepts can be useful for the beginners. The implemented code solutions for different problems are a good set of exercises for the students. Moreover, these codes can serve as a baseline when students face bigger projects.

Supplemental Resources

This book is accompanied by a set of IPython Notebooks containing all the codes necessary to solve the practical cases of the book. The Notebooks can be found on the following GitHub repository: <https://github.com/DataScienceUB/introduction-datascience-python-book>.

Acknowledgements

We acknowledge all the contributing authors: J. Vitrià, E. Puertas, P. Radeva, O. Pujol, S. Escalera, L. Garrido, and F. Dantí.

Barcelona, Spain

Laura Igual
Santi Seguí



<http://www.springer.com/978-3-319-50016-4>

Introduction to Data Science

A Python Approach to Concepts, Techniques and
Applications

Igual, L.; Seguí, S.

2017, XIV, 218 p. 73 illus., 67 illus. in color., Softcover

ISBN: 978-3-319-50016-4