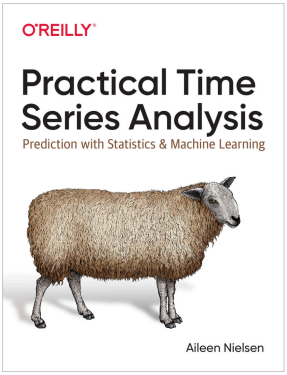
ebookbe.com

PRACTICAL TIME SERIES ANALYSIS PREDICTION WITH STATISTICS AND MACHINE LEARNING (AILEEN NIELSEN)



Scan Qr Code To Access Full Version

Full Version On:

ebookbe.com

Get instant access to the complete digital publication, supplementary materials, and full unabridged content. Scan the QR code above or visit our website.

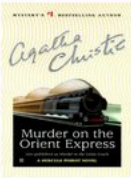
Recommended Course Materials & Supplementary Reference Resources

Please note: Below are recommended reference materials that accompany this subject. Scan the QR code next to any title to view syllabus details.



Modern Reloading (Richard Lee)

Storeys Guide to Keeping Honey Bees Honey Production, Pollination, Bee Health (Richard E.



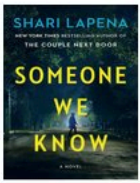
Murder on the Orient Express (Agatha Christie) (1)



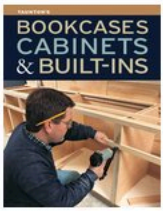
The Bravest Kid (Sally Rippin)



Sustainable Facades Design Methods for High-Performance Building Envelopes (Ajla



Someone We Know (Shari Lapena)



Taunton's Bookcases, Cabinets Built-Ins (coll.)



That Time I Got Drunk And Saved A Demon (Kimberly Lemming)

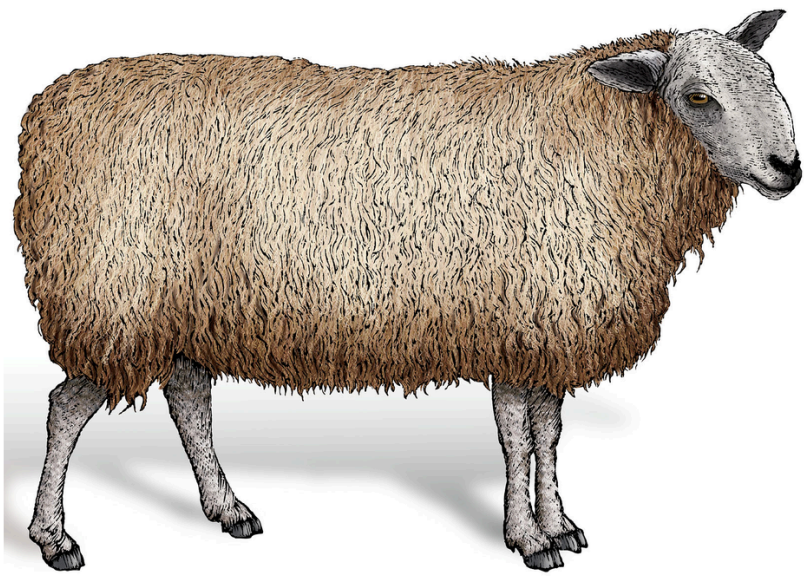


Symmetry and Spectroscopy An Introduction to Vibrational and Electronic Spectroscopy (Daniel C.

O'REILLY®

Practical Time Series Analysis

Prediction with Statistics & Machine Learning



Aileen Nielsen

Practical Time Series Analysis

Prediction with Statistics and Machine Learning

Aileen Nielsen

ght © 2020 Aileen Nielsen. All rights reserved.

in the United States of America.

hed by O'Reilly Media, Inc., 1005 Gravenstein Highway North, opol, CA 95472.

ly books may be purchased for educational, business, or sales tional use. Online editions are also available for most titles (*http:// com*). For more information, contact our corporate/institutional sal ment: 800-998-9938 or *corporate@oreilly.com*.

ment Editor: Jeff Bleiel	Indexer: WordCo Indexing Services, Inc.
tions Editor: Jonathan Hassell	Interior Designer: David Futato
tion Editor: Katherine Tozer	Cover Designer: Karen Montgomery
itor: Rachel Monaghan	Illustrator: Rebecca Demarest
ader: Sonia Saruba	

October 2019: First Edition

on History for the First Edition

2019-09-20: First Release

p://oreilly.com/catalog/errata.csp?isbn=9781492041658 for release .

Reilly logo is a registered trademark of O'Reilly Media, Inc. *Practic eries Analysis*, the cover image, and related trade dress are tradema ly Media, Inc.

ews expressed in this work are those of the author, and do not repr

tions contained in this work is at your own risk. If any code sampl
echnology this work contains or describes is subject to open source
s or the intellectual property rights of others, it is your responsibili
that your use thereof complies with such licenses and/or rights.

492-04165-8

er, stock markets, and heartbeats. They all form time series. If you' ted in diverse data and forecasting the future, you're interested in t analysis.

me to *Practical Time Series Analysis*! If you picked this book up, you' ly already noticed that time series data is everywhere. Time series increasingly ubiquitous and important as the big data ecosystem s. For better or worse, sensors and tracking mechanisms are every a result there are unprecedented amounts of high-quality time seri vailable. Time series are uniquely interesting because they can addr ns of causality, trends, and the likelihood of future outcomes. This ke you through the major techniques that are most commonly appl ries to address such questions.

eries data spans a wide range of disciplines and use cases. It can be ng from customer purchase histories to conductance measurements lectronic system to digital recordings of human language. One poi s throughout the book is that time series analysis applies to a surpri set of data. Any data that has an ordered axis can be analyzed wit ries methods, even if that ordered axis is not time per se. Tradition ries data, such as stock data and weather patterns, can be analyze ries methods, but so can quirky data sets such as spectrographs of the "time" axis is actually an axis of frequency. Time series are here.

Should Read This Book

are two kinds of intended readers for this book. The first and larger ry of reader is that of a data scientist who has rarely worked with t data. This person could be an industry veteran or a junior analyst. xperienced data analyst can skim the introductory conceptual area hapter but will still benefit from this book's discussions about best es as well as pitfalls of working with time series data. A newer dat

u be currently coding in your professional life. For such a reader, t s useful to point out opportunities for your organization to use time analysis even if it is not currently practiced in-house. This book wil ou to new kinds of questions and analyses your organization can a our existing data resources.

ted Background

espect to coding, you should have some familiarity with R and Pyt ally with certain fundamental packages (in Python: NumPy, Pandas -learn; and in R: `data.table`). The code samples should be readab ithout all the background, but in that case you may need to take a to familiarize yourself with these packages. This is most likely the espect to R's data.table, an underused but highly performant data fr e that has fantastic time functionality.

ases, I have provided brief overviews of the related packages, som le code, and descriptions of what the code does. I also point the rea more complete overviews of the most used packages.

espect to statistics and machine learning, you should have some rity with:

uctory statistics

as such as variance, correlation, and probability distributions

ne learning

stering and decision trees

networks

at they are and how they are trained

ese cases, I provide a brief overview of such concepts within the tex initiated should read more deeply about them before continuing wi hapters. For most topics, I provide links to recommended free onli ces for brief tutorials on the fundamentals of a given topic or techn

ime series is an important aspect of data analysis but one that is no in the standard data science toolkit. This is unfortunate both becau ries data is increasingly available and also because it answers ques oss-sectional data cannot. An analyst who does not know fundame ries analysis is not making the most of their data. I hoped that this fill an existing and important void.

, when I started writing this book, I knew of no centralized overvi st important aspects of time series analysis from a modern data sci ctive. There are many excellent resources available for traditional t analysis, most notably in the form of classic textbooks on statistical analysis. There are also many excellent individual blog posts on bot onal statistical methods and on machine learning or neural networ ches to time series. However, I could not identify a single centraliz ce to outline all these topics and relate them to one another. The g ok is to provide that resource: a broad, modern, and practical over series analysis covering the full pipeline for time series data and ing. Again, I hoped that this book could fill an existing and importa

time series is an interesting topic with quirky data concerns. Probl ted with data leakage, lookahead, and causality are particularly fu time series perspective, as are many techniques that apply uniquel rdered along some kind of time axis. Surveying these topics broadl a way to catalog them was another motivation to write this book.

gating This Book

ook is organized roughly as follows:

y

apter 1 presents a history of time series forecasting, all the way fro ient Greeks to modern times. This puts our learning in context so erstand what a traditional and storied discipline we study in this b

ut the data

ortant in most data pipelines. Data identification and cleaning rep
rge portion of the work done by most time series analysts.

s, models, models

apters 6, 7, 8, 9, and 10 cover a variety of modeling techniques tha
used for time series analysis. We start with two chapters on statisti
thods, covering standard statistical models, such as ARIMA and Ba
te space models. We then apply more recently developed methods,
machine learning and neural network, to time series data, highlight
challenges of data processing and data layout when time series dat
d for fitting models that are not inherently time aware, such as dec
es.

odeling considerations

apters 11 and 12 cover accuracy metrics and performance consider
pectively, to provide some guidance as to what you should conside
have performed your first iterations of time series modeling.

orld use cases

apters 13, 14, and 15 provide case studies from healthcare, finance,
ernment data, respectively.

ents on recent happenings

apters 16 and 17 briefly cover recent time series developments and
dictions for the future. Chapter 16 is an overview of a variety of
omated time series packages, some open source and developed as a
demic effort, and some coming from large tech companies. These t
under active development as efforts ramp up to improve time serie
ecasting at scale via automated processes. Chapter 17 discusses so
dictions for the future of time series analysis as the big data ecosys
ws and as we learn more about how big data can assist time series
lysis.

eral, I recommend reading through a chapter before attempting to
e code. There are usually a few new concepts introduced in each
r, and it can be helpful to give these some attention before turning

he book is written so that it is most sensibly read from beginning t
ater chapters referring to concepts covered earlier in the book), bu
I purposely kept the chapters as self-contained as possible to enabl
xperienced readers to jump around as convenient.

e Resources

ly hosts a GitHub repository for this book. The repository includes
code discussed in this book in a form where you can run it on data
are also provided in the repository. In some cases, variable names,
ting, and such will not be identical to that produced in the book, b
be easily linked (for example, in some cases variable names were
ned in the book due to formatting constraints).

onally, should you enjoy video presentations, I have two online tut
ng some of the content of this book with a Python focus. If you wan
ement this book with a video presentation, consider the following
ces:

Time Series Analysis (SciPy 2016)

Modern Time Series Analysis (SciPy 2019)

entions Used in This Book

llowing typographical conventions are used in this book:

icates new terms, URLs, email addresses, filenames, and file extens

`nt width`

d for program listings, as well as within paragraphs to refer to pro
ments such as variable or function names, databases, data types,
ironment variables, statements, and keywords.

ws text that should be replaced with user-supplied values or by val
ermined by context.

> **Tip**
>
> This element signifies a tip or suggestion.

> **Note**
>
> This element signifies a general note.

> **Warning**
>
> This element indicates a warning or caution.

g Code Examples

mental material (code examples, exercises, etc.) is available for
oad at *https://github.com/PracticalTimeSeriesAnalysis/BookRepo*.

ook is here to help you get your job done. In general, if example co
with this book, you may use it in your programs and documentati
not need to contact us for permission unless you're reproducing a
cant portion of the code. For example, writing a program that uses
l chunks of code from this book does not require permission. Sellin