



Python Tips for Data Scientist

Release 1.00

Wenqiang Feng

February 22, 2019

CONTENTS

| | | |
|----------|----------------------------------------|----------|
| 1 | Preface | 3 |
| 1.1 | About | 3 |
| 1.2 | Motivation for this tutorial | 4 |
| 1.3 | Feedback and suggestions | 5 |



Welcome to my **Python Tips for Data Scientist** notes! In those notes, you will learn some useful tips for Data Scientist daily work. The PDF version can be downloaded from [HERE](#).

PREFACE

Chinese proverb

Good tools are prerequisite to the successful execution of a job. – old Chinese proverb

1.1 About

1.1.1 About this tutorial

This document is a summary of my valueable experiences in using Python documentation Sphinx with Github webpage. The PDF version can be downloaded from [HERE](#). **You may download and distribute it. Please be aware, however, that the note contains typos as well as inaccurate or incorrect description.**

In this repository, I try to use the detailed demo code and examples to show how to use Sphinx to generate the .html and .pdf documents and how to hookup them automatically on Github. If you find your work wasn't cited in this note, please feel free to let me know.

Although I am by no means a python programming and Sphinx expert, I decided that it would be useful for me to share what I learned about Sphinx in the form of easy tutorials with detailed example. I hope those tutorials will be a valuable tool for your studies.

The tutorials assume that the reader has a preliminary knowledge of python programing, LaTeX and Linux. And this document is generated automatically by using [sphinx](#).

1.1.2 About the authors

- **Wenqiang Feng**
 - Data Scientist and PhD in Mathematics
 - University of Tennessee at Knoxville

– Email: von198@gmail.com

- **Biography**

Wenqiang Feng is Data Scientist within DST's Applied Analytics Group. Dr. Feng's responsibilities include providing DST clients with access to cutting-edge skills and technologies, including Big Data analytic solutions, advanced analytic and data enhancement techniques and modeling.

Dr. Feng has deep analytic expertise in data mining, analytic systems, machine learning algorithms, business intelligence, and applying Big Data tools to strategically solve industry problems in a cross-functional business. Before joining DST, Dr. Feng was an IMA Data Science Fellow at The Institute for Mathematics and its Applications (IMA) at the University of Minnesota. While there, he helped startup companies make marketing decisions based on deep predictive analytics.

Dr. Feng graduated from University of Tennessee, Knoxville, with Ph.D. in Computational Mathematics and Master's degree in Statistics. He also holds Master's degree in Computational Mathematics from Missouri University of Science and Technology (MST) and Master's degree in Applied Mathematics from the University of Science and Technology of China (USTC).

- **Declaration**

The work of Wenqiang Feng was supported by the IMA, while working at IMA. However, any opinion, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the IMA, UTK and DST.

1.2 Motivation for this tutorial

Sphinx is an awesome Python documentation package, and it has excellent facilities for the documentation of software projects in a range of languages. I have been using Sphinx for almost 4 years. I was impressed and attracted by Sphinx in the first using. And I found that:

- It supports **several popular output formats**: HTML (including Windows HTML Help), LaTeX (for printable PDF versions), ePub, Texinfo, manual pages, plain text.
- It has **easy publishing routes**: Github.
- It has **extensive cross-references**: semantic markup and automatic links for functions, classes, citations, glossary terms and similar pieces of information
- It has **hierarchical structure**: easy definition of a document tree, with automatic links to siblings, parents and children.
- It has **automatic indices**: general index as well as a language-specific module indices
- It has awesome **code handling**: automatic highlighting using the Pygments highlighter

- It has abundant **extensions**: automatic testing of code snippets, inclusion of docstrings from Python modules (API docs), and more
- It has abundant **contributed extensions**: more than 50 extensions contributed by users in a second repository; most of them installable from PyPI

1.3 Feedback and suggestions

Your comments and suggestions are highly appreciated. I am more than happy to receive corrections, suggestions or feedbacks through email (Wenqiang Feng: von198@gmail.com) for improvements.