

# 《Python数据科学入门》阅读计划

——图灵后端群阅读计划（第2期）

领读人：张旱文

## 本书特色

- 让读者亲切体会到不同类型文本数据(csv、json、自然语言中的文本)的获取、清洗、组织和可视化
- 使用NumPy和Pandas模块处理数值数据
- 实战分别用MySQL、MongoDB数据库进行配置、填充、查询数据
- 基于网络和非网络的数据,创建网络、网络度量和分析网络
- 概率与统计以及机器学习的相关基本概念

## 适合读者

- 刚入门的数据科学专业人员
- 数据科学教员
- Python语言爱好者
- 研究生和本科生
- 想了解大数据分析和机器学习的兴趣爱好者
- 想拥有一本参考手册来帮助记住所有Python函数及参数的开发人员

总阅读时长：3~4 周

每天阅读用时：1~2小时（建议多做练习）

答疑时间安排：每周一次，周六 20:00~22:00

图灵社区本书网址：<http://www.ituring.com.cn/book/1919>

图灵阅读计划网址：<https://github.com/BetterTuring/turingWeChatGroups>

## 读前须知

---

- 书中代码你可以从Github上获取，[点击获取](#)。如果你觉得代码有不妥的地方，可以提出issue，或者将你自己认为比较好的代码进行Pull requests，我会对有价值的issue、Pull requests非常感谢，同时发放一些奖励。
- 为方便记录大家学习的笔记，同时帮助我们学习用Markdown来记录笔记，当然你也可以选择用其它的方式来记录自己学习的笔记。我在GitHub上创建了一个仓库，大家可以把每天的学习笔记进行PR。可

以让更多的后来读者看到你们的贡献，这将是一件多么伟大的事情。有贡献的读者将会获取对应的奖励，我们要尊重知识的价值。

1. 仓库地址：<https://github.com/zhanwen/PythonDataScience>
  2. 目录已创建完成。为了方便记录和区分，大家在PR的时候，建议文件命名规则为：**作者姓名英文缩写+笔记文件名**。在`note`目录中已有示例。
- 另外，我会找一些其他资料中跟本书知识点相关的笔记，让大家一起来练习，以便更好地掌握本书知识。
  - 如果读者有任何不清楚的地方或者对一些技术(Markdown、git)搞不明白，都可以在[这里](#)提出Issue，也可以通过最下面的方式与我联系，我将会用所有知识替你们解忧^\_^。

## 阅读规划

---

### 第一部分（1~2章）

阅读时长：1周之内

#### 基础部分

- 对数据科学有个初步的认识
  1. 数据分析步骤
  2. 数据的获取途径
  3. 报告的结构

#### 重点部分（实战）

- Python 的使用，对没有 Python 编程经验的，要更多的练习
  1. 几本的字符串函数使用
  2. Python 中的数据结构
  3. Python 中的文件使用
  4. 正则表达式
  5. Pickling 和 Unpickling 数据

### 第二部分（3~5章）

阅读时长：1~1.5周

#### 基础部分

- Python 的使用，巩固第二章所学的知识
- 了解文本数据的格式(csv, html, json)
- MySQL、MongoDB 的概念理解

## 重点部分（理解与实战）

- 使用 Python 去处理文本数据
- Mysql 的命令行操作以及使用 Python 来操作 Mysql
- MongoDB 的安装，使用 Python 来操作 MongoDB 。
- 数组的索引和切片，聚合与排序
- 数组的保存和读取
- 如何合成正弦波

## 第三部分（6~7章）

阅读时长：1周之内

### 基础部分

- 理解 pandas 的数据结构
- pandas 模块里的 series、frame 的使用
- 理解网络数据的概念

## 重点部分（理解与实战）

- 使用 pandas 处理一些常见的问题
  1. 数据重塑
  2. 处理缺失的数据
  3. 组合数据
  4. 数据的排序和描述
  5. 数据之间的转换
  6. 文件的读写
- 基于网络的和非网络的数据创建网络
  1. 网络度量
  2. 网络分析序列

## 第四部分（8~10章）

阅读时长：1~1.5周

## 基础部分

- 了解可视化工具，绘图类型
- 概率与统计的一些基本概念
- 机器学习的基础知识

## 重点部分（理解与实战）

- 使用 **Pyplot** 进行绘图，并可以进一步对绘图进行装饰
- 使用 **pandas** 绘图
- 以 Python 的方式完成统计
- 线性回归你拟合
- k 均值聚类实现数据分组
- 随机决策森林

## 其他建议

---

- 由于每个人学习的方式不同以及读书进度不同，大家可以在建议阅读时长上进行调整
- 对理解不透的知识，我们可以在读者群里一起讨论，或者通过下面的方式与我联系，我将会尽可能的帮助
- [1106002609@qq.com](mailto:1106002609@qq.com)（张旱文）