

# 《Python数据分析基础》 阅读计划

——大数据与机器学习群阅读计划（第1期）

领读人：天予

## 本书特色

- 零编程基础要求，任何人皆可上手
- 重在代码实现，不拘泥于理论

## 适合读者

- 经常做重复性工作（比如 Excel）的办公者
- 对数据可视化感兴趣的读者（入门级）

## 阅读建议

- 本书重在实现，任何人都可以学习，部分章节可以酌情跳过
- 部分章节背后理论复杂，感兴趣的可以自行探索

阅读总时长：3周

阅读日时长：2小时

答疑时间：每周日晚 20:00—22:00

图灵社区本书网址：<http://www.ituring.com.cn/book/1912>

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

## 阅读规划

第一周：第1~2章

### 重点

1. Python基础的学习或复习

2. 筛选csv文件中需要的部分以及对多个文件进行合并及处理

## 难点

- 零编程基础看第1章可能难以理解，不如快速浏览一遍 [《Python编程：入门与实践》](#) 的基础知识部分
- 对 csv 文件的基础 Python 处理以及使用 Pandas 的处理，有需要可以继续自行学习

## 第二周：第3~5章

### 重点

1. 处理Excel工作表，工作簿，及多个工作簿
2. Python 与 MySQL 的结合使用
3. 大批量自动处理csv以及文本文件的一些常用任务

### 难点

- 如果办公 Excel 常用可以重点研究之，以及自行探索
- 对 MySQL 需要一点了解，视需要可以酌情跳过

## 第三周：第6~9章

### 重点

1. Matplotlib下的五种常用图形绘制
2. 对一个实例的直方图、t检验、相关性、线性回归以及预测
3. 各环境下的自动运行脚本方法
4. 对 Python 的一些扩展

### 难点

- 描述性统计难点不在于如何绘图，而在于如何看懂并阐释它
- 某些你可能不熟悉的统计方法，例如logistic回归
- 各种 Python 扩展包的使用以及数据结构的深入研究（如果你有兴趣继续下去）