

# 从Kaggle比赛深入数据科学

## 课程准备

这门课程以Python作为教学语言，以Jupyter Notebook作为编程交互环境。在上课前，请同学们安装好Anaconda，Anaconda中包括Python和Python中的自带常用数据科学工具，也包括Jupyter Notebook。Anaconda的详细安装方法请参考：[嫌Python安装太麻烦？你该试试Anaconda](#)

本文中也简要介绍Anaconda的安装以及Jupyter Notebook的使用，基本步骤如下：

### 1. 下载Anaconda

到[Anaconda的官网上](#)，选择一个合适的版本下载。建议下载Python 3.7 version。Python 3.7原生支持utf-8，对中文的支持比较好，不用复杂的转码。

### 2. 安装Anaconda

按照安装程序的提示一步一步的安装。

### 3. 运行Jupyter Notebook

到Windows自带的命令提示符下（开始菜单→Windows系统→命令提示符），输入 `jupyter notebook`，回车，会在浏览器中弹出一个页面，这个页面就是Jupyter Notebook编辑器，可以在里面写代码啦。（苹果系统中是在终端中输入jupyter notebook，回车。）

### 4. 创建代码文件

现在在界面中看到是 `c:\users\用户名\` 下面的一些文件夹。选择一个文件夹，作为你保存代码的地方。比如Documents（就是文件资源管理器里面的“文档”文件夹），点击进入。右上角有一个“New”的按钮，点击那个按钮，然后选择“Python 3”（如果之前下载的是python 2.7版本的anaconda，这里就选择“Python 2”。现在就可以写代码啦。

### 5. 编辑器简介

Jupyter Notebook中的代码是一块一块的，点击左上角的加号，就可以增加一个代码块。选中一个代码块，写好代码以后，点击上面的三角形播放符号，就运行这一块的代码，输出的结果会直接显示在这个代码块下面，马上可以看见结果。这种编程模式称为交互式编程，是最适合数据科学的一种编程模式。点击左上角的保存符号，就保存了代码文件（包括所有输出的结果）。其他的功能可以慢慢摸索。

**注：**如果已经对Python环境配置非常熟悉，不必按照以上方法配置Python环境。使用自己喜欢的环境管理工具和编辑器就可以