

# 夏强

- Github: <https://github.com/ShawnXiha>
- Email: huakaijianwo2009@163.com
- Phone: 18679188050
- Location: 下城区, 杭州, 浙江

## • 摘要

我是一名刚毕业转行数据分析的食品科学研究生。在导师团队我处理了大量的实验数据,发现用spss处理数据非常低效繁琐,所以到网上寻找更好的方法。在见识了python和R在处理数据时的强大功能之后,我仿佛打开了新世界的大门,决定做数据分析师。

我开始在网上自学编程和数据处理。现在,我从Udacity纳米学完毕业了,我可以熟练的使用python、R和SQL处理分析数据,使用机器学习探寻数据中的规律,并将数据可视化。

## • 技能:

- 语言: python,R,Javascript,HTML5,CSS3,
- 库: NumPy,Pandas,Matplotlib,scikit-learn,d3.js,dimple.js, ggplot2
- 其它: git/github,SQL,MongoDB

## 项目经验

[https://github.com/ShawnXiha/Udacity\\_Data\\_Analyst\\_Nanodegree](https://github.com/ShawnXiha/Udacity_Data_Analyst_Nanodegree)

### 1. 调查数据集

- 使用Pandas整理清洗数据集,使用Matplotlib作图分析数据分布规律,及变量间关系
- 使用scipy对各因素对结果影响进行假设检验

### 2. 探索并总结数据

- 使用ggplot2数据可视化对数据进行单变量,双变量,多变量分析
- 使用GGally探索变量之间的相关性

### 3. 从安然的邮件中识别舞弊

- 发现并移除数据集中的离群值,异常值,使用主成分分析及特征选择对机器学习的数据进行预处理
- 选取性能度量指标,采用交叉验证选择最优的机器学习算法并调节参数

#### 4. 处理 OpenStreetMap 数据

- 解析上海市开放地图数据并发现其中存在的问题
- 将地图数据中英文混合，中文及拼音混合，英文的街道名修改成中文名
- 将清洗好的地图数据导入MongoDB，并用MongoDB分析数据

#### 5. 高效数据可视化

- 发现数据规律，并使用d3.js及dimple.js来展示分享
- 接收反馈，并根据反馈意见提高数据可视化的效果

#### 6. A/B 测试

- 使用A/B测试分析，Udacity如果在注册学习前提示每周最少需要投入的学习时间能否可以在减少免费试学学生数量的同时不减少付费学生数量，并判断这个决定可不可取。
- 设计一个后续试验来减少免费试学后取消付费学生的人数。

---

## 教育

**Udacity** 数据分析师纳米学位 2016.6~2016.9

南昌大学 食品科学硕士 2013.9~2016.6

## 其他技能：

设计、计划、安排、实施实验并分析处理数据写成文章

做ppt，演讲，英语6级，番茄工作法，