

# 数据分析师 SVIP 特训班

## 张可人 学习报告



2019.3 Version 1.0

### 学员信息

姓名：张可人

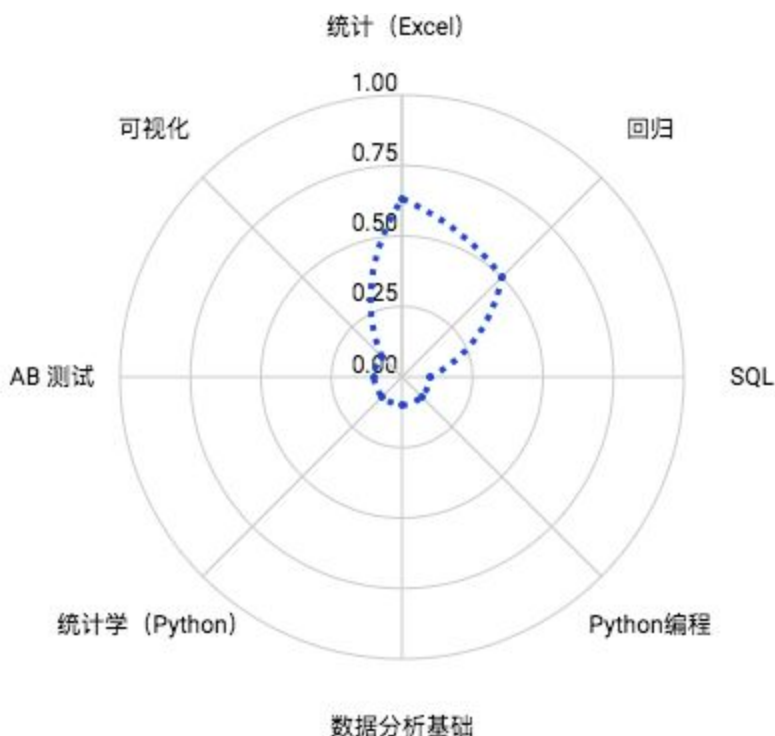
年龄：27

工作背景：零售行业从业4年多，最开始做管培生。目前做咨询相关工作有1年多，涉及对电商的数据分析，但目前都以excel应用较多。

学习经历：UIUC经济学背景毕业，没有编程和算法基础

### 技术能力背景分析

张可人同学于3月19日完成技术能力测试，涉及统计学（Excel）、回归、SQL、Python编程、数据分析基础、统计学（Python）、AB 测试和可视化 8个纬度。经过测试，你的各纬度能力分布如下：



根据你的测试结果，Udacity 服务团队对你的建议如下：

- 统计学（Excel）：你可能会需要统计学的知识（比如：分布、描述统计学、电子表格基本操作）来进行了解业务的现实状况。  
目前你的掌握程度为：良好
- 回归：你可能会需要掌握回归和建模的能力（比如：线性回归基础、线性回归的前提）来进行商业预测  
目前你的掌握程度为：有待加强
- SQL：你可能会需要利用 SQL 从公司数据库中提取需要分析的数据，作为后续数据分析的基础，目前你  
目前你的掌握的程度为：有待加强
- Python编程：你可能会需要 Python 编程能力来对大规模的数据进行清洗和分析。  
目前你的掌握程度为：有待加强
- 数据分析基础：你的目标是成为一名数据分析师，你需要掌握数据分析（提出问题、数据整理、数据处理、发现洞察）的能力。  
目前你的掌握程度为：有待加强
- 统计学（Python）和 A/B 测试：你可能会需要概率计算、假设检验、设计A/B 测试、并评估结果的能力来辅佐商业决策，  
目前你的掌握程度为：有待加强
- 可视化：你可能需要了解可视化的设计原则以完成合适的图表呈现，帮助公司业务方和高层了解运营数据，为决策提供数据支持。  
目前你的掌握程度为：有待加强

## 学习目标和求职意向

Udacity服务团队与张可人同学于3月21日进行了视频面聊，了解你的学习目标如下：

晋升意向
<ol style="list-style-type: none"> <li>1. 是否有求职/转岗/晋升意向：转岗</li> <li>2. 目标行业Top3：涉及数字化的咨询行业或其他行业涉及数字化的部门</li> <li>3. 目标城市Top2：继续在上海</li> <li>4. 目标求职/转岗/晋升时间：19年年底</li> </ol>
时间投入
<p>6个月内的时间投入：工作比较弹性，出差较多，一周可以保证15h</p> <p>6个月内的特殊事项：无，可全身心投入学习</p>

## 职业发展建议

## 学习/工作状态

日常的工作职责：快消品行业咨询岗位，涉及到很多数据分析的工作

学习动因：工作中喜欢用数据讲故事，目前主要用excel做数据分析，感觉说服力较为薄弱，希望有更有力的数据分析技能作为支撑

## 自我认知

自身优势（最希望他人记住的品质）：拥有很强的学习热情和工作上面的需求

自身较为薄弱的环节（与自我目标相比）：编程和算法没有接触过，工具和知识方面比较欠缺，缺少相关行业人脉

## 职业发展“痛点”

求职/晋升目标：数字化想的咨询行业

求职/晋升动机：强

求职/晋升策略：一般

其他：对于目标岗位比较清晰，但具体行业未确定

## 期许

下一步行动计划：

- 全面提升数据分析的硬性能力；
- 参与毕业生和行业导师分享，拓展行业知识面；
- 再次进行 1 对 1 职业咨询，明确下一步行动计划；

## 推荐课程及个性化安排

根据你的技术能力基础、学习目标、求职意向和时间投入，Udacity 中国数据科学服务团队为你推荐课程模块如下，你可以查看对应的推荐掌握程度和推荐学习方式。你的助教同样会根据该推荐的课程模块，帮助你规划导学内容，监督你的学习进度，帮助你复盘所学内容。

推荐掌握程度和推荐学习方式的细则对应如下：

推荐掌握程度		推荐学习方式	
了解	基本上只停留在听说过，见过，了解过的阶段；	简单浏览	查看视频内容，了解基本术语，不做项目要求

熟悉	可能是接触过、用过、经历过，具体理论或操作有认知，但不一定能独立工作或独立应用；	精进学习	曾经见过接触过，但仅停留在了解阶段，建议系统学习
精通	基本上可以独立处理这一方面的问题，有自己的见解；	实战应用	与未来规划契合度相当高，且具有强实战性，建议多应用多演练
/	/	*回顾复盘	系统学习过，能力基本满足要求，需要定期回顾复盘

个性化课程推荐如下：

模块	课程名称	学习目标	项目	推荐掌握程度	推荐学习方式
数据处理	数据处理	基于数据分析经典案例，初步了解使用电子表格和 SQL 做数据分析	观察数据可视化面板完成分析报告	精通	实战应用
描述统计学	描述统计学 1	学习数据类型，集中趋势的测量方法，以及基本的数学符号	分析问卷数据	精通	实战应用
	描述统计学 2	学习针对定量数据的常用可视化方法，离散程度的测量方法，以及描述统计学和推论统计学的差异			
	Excel：开始	学习 Excel 软件的基本功能 学习单元格引用以及菜单的快捷键			
	Excel：数据清洗	学习筛选和排序数据 学习使用文本和数学公式 学习如何拆分列以及去除重复项			
	Excel：分析数据	学习使用聚合函数和条件函数来汇总数据 学习使用数据透视表和 lookup 查询函数			
	Excel：数据可视化	学习为定量数据和分类数据创建可视化 学习创建饼图、条形图、折线图、散点图、直方图和箱线图 学习制作专业的演示 PPT			
用数据分析解决商业问题	分析问题解决框架	学习行业流行的商业分析问题解决框架	预测商业决策带来的收益增长	精通	实战应用
	选择合适的分析方法	学习针对不同商业问题的分析方法，包括：预测性分析和非预测性分析，数据丰富和数据有限，数值问题、分类问题、A/B 测试和用户群体划分等			
	用 Excel 构建线性回归模型	学习线性回归模型背后的数学知识 学习用 Excel 搭建简单的线性回归模型 学习用 Excel 搭建多元线性回归模型			
数据分析与 SQL	基础 SQL	编写基本 SQL 命令，如 SELECT、FROM 和 WHERE，以及相应的逻辑运算符	基于零售数据挖掘业务特征	精通	实战应用
	SQL 连接（SQL Joins）	在 SQL 中编写 Join 命令，从而整合多个数据库中的数据解决复杂的商业			

		问题			
	SQL 聚合函数	编写SQL基本聚合函数，包括 COUNT, SUM, MIN 和MAX 编写 CASE 和 DATE 函数，并处理 NULL 值			
	SQL 高级数据查询	使用 CREATE TABLE, INSERT INTO 和UPDATE 运算符及其他语句编写数据库 使用窗口函数和子查询添加查询步骤 使用文件学习新函数并处理复杂任务			
Python 入门	为何要学习 Python 编程	了解我们为什么要学习编程 了解使用 Python 编程的独特特点	探索美国共享单车数据	熟悉	精进学习
	数据类型和运算符	了解并使用以下数据类型：整型，浮点型，布尔型，字符串，类别，元组，集合，字典 了解并使用以下运算符：算数运算符，赋值运算符，比较运算符，逻辑运算符，成员运算符，恒等运算符			
	控制流	编写 if 语句表示条件判断 编写 for 和 while 循环语句表示重复			
	函数	将编码组织成函数 了解变量作用域 文档和注释			
	脚本编写	在本地计算机上编写和运行脚本 处理用户的原始输入 读取和写入文件，处理错误，并导入本地脚本 使用来自 Python 的标准库和第三方模块 使用在线资源来帮助解决问题			
	NumPy	学习 NumPy 的基础知识 使用 NumPy 创建和操作数组			
	Pandas	学习 Pandas 的基础知识 使用 Pandas 创建和操作 Series 和 Dataframe			
数据分析入门	数据分析流程	了解数据分析流程的主要步骤 运用 Python 和 Pandas 处理多个数据集	探索数据集	熟悉	精进学习
	Pandas 和 Numpy：案例分析1	对一个数据集进行完整的数据分析 学习使用 NumPy 和 Pandas 进行数据的整理、探索、分析及可视化处理			
	Pandas 和 Numpy：案例分析2	对一个数据集进行完整的数据分析 深入学习使用 NumPy 和 Pandas 进行数据的整理、探索、分析及可视化处理			
统计学 1 - 推论统计学	概率和条件概率	了解概率基础知识 学习正态分布和二项分布 理解条件概率和贝叶斯规则	测试心理学现象	熟悉	精进学习
	抽样分布和中心极限定理	了解抽样分布以及编程实现 了解中心极限定理 自助法 (bootstrap) 抽样			
	假设检验	了解置信区间和 p 值 假设检验、得出结论和常见的错误类型			

数据清洗	数据清洗入门	了解数据整理流程的各个步骤（收集、评估和清洗） 利用基本的数据收集、评估和清洗代码来整理从 Kaggle 下载的 CSV 文件	清洗和分析数据	了解	简单浏览
	收集数据	收集不同来源的数据，包括收集文件、以编程方式下载文件、网络抽取数据和访问API数据等 将不同文件格式的数据导入 Pandas, 包括平面文件（如 TSV）、HTML 文件、TXT 文件和 JSON 文件 将收集到的数据储存在 PostgreSQL 数据库中			
	评估数据	使用 Pandas 以及编程的方式直观地评估数据 区分脏数据（内容或质量问题）和乱数据（结构或整齐程度问题） 辨别数据质量问题并用矩阵进行分类：有效性、准确性、完整性、一致性和统一性			
	清洗数据	了解数据清洗的各个步骤（定义、编码和检验） 使用 Python 和 Pandas 清洗数据 使用 Python 以直观的及编程的方式检验清洗代码			
探索性数据分析	什么是 EDA	明确并了解探索性数据分析（EDA）的重要性	用 R 探索和总结数据	了解	简单浏览
	R 基本知识	安装 RStudio 软件及程序包 编写基本 R 语言脚本以检测数据集			
	探索一个变量	量化并可视化数据集中的单个变量 创建直方图和箱线图 变换变量 检查并识别可视化中的得失			
	探索两个变量	合理运用相关技巧探索数据集中任意两个变量间的关系 创建散点图 计算相关性 探讨条件均值			
	探索多个变量	重塑数据框，并运用色彩、形状等美学元素揭示信息			
	钻石和价格预测	利用预测模型为钻石定价			
通过数据讲故事	数据可视化基本原理	了解数据可视化的重要性 了解不同数据类型如何进行可视化编码	创建一个 Tableau 故事	精通	实战应用
	设计原则	根据数据特点选择最有效的图表 有效运用色彩、形状、大小等元素			
	用TABLEAU创建可视化	熟悉使用 Tableau 基本功能，如图表、过滤器、分层结构等 创建 Tableau 计算字段			
	用TABLEAU讲故事	创建 Tableau 仪表盘和故事，展示有效的数据可视化			
统计学 2 - A/B 测试与回归	A/B 测试	案例学习：A/B 测试是如何在线上教育公司 Udacity 所使用的	分析实验结果	了解	简单浏览

	回归	构建单元/多元线性回归模型，理解自变量和因变量的关系 利用单元/多元线性回归的结论做出预测			
	分类	学习使用 Python 构建逻辑回归模型 学习解读结果以及衡量模型好坏			
数据分析 职业发展	个人简历的创建和优化	学习更好的展示自己的简历，呈现项目经历	个人简历的创建和优化	精通	实战应用
	行为导向面试	学习在面试中做更好的行为导向面试	行为导向面试	精通	实战应用
	技术模拟面试	学习在面试中更好的回答技术问题	技术模拟面试	精通	实战应用
监督学习	线性回归	了解分类与回归的区别，学习如何使用线性回归来做预测	寻找慈善机构的捐助者	了解	简单浏览
	感知器算法	学习神经网络中的感知器，以及如何使用它进行分类			
	决策树	学习决策树，并使用决策树探索泰坦尼克号乘客存活模型			
	朴素贝叶斯	学习朴素贝叶斯原理，并构建垃圾邮件分类器			
	支持向量机	学习如何训练支持向量机以线性分离数据； 使用核方法在非线性可分的数据上来训练 SVMs			
	集成方法	通过 boosting 提升传统方法； Adaboost			
	自我评估：监督学习	监督学习相关的测试题			
非监督学习	聚类	学习如何聚类算法，并尝试使用 k-means 对数据进行聚类	创建客户细分	了解	简单浏览
	聚类迷你项目	使用 k-means 对电影评分进行聚类			
	层次聚类法与密度聚类	学习单连接聚类法和层次聚类法，DBSCAN			
	高斯混合模型与聚类验证	学习高斯混合模型及相关示例			
	特征缩放	通过案例学习特征缩放			
	PCA（主成分分析）	了解降维的作用，并学习 PCA 的原理和使用场景			
	PCA 迷你项目	使用特征脸方法和 SVM 进行脸部识别			
	随机投影与 ICA	学习随机投影与独立成分分析，并通过 Lab 学习如何应用这些方法			
	非监督学习自我评估	非监督学习相关的测试题			
神经网络	深度学习简介	欢迎学习此课程 应用深度学习 Anaconda Jupyter Notebooks 矩阵数学和Numpy复习	预测共享单车使用情况	了解	简单浏览
	神经网络	神经网络简介 实现梯度下降 训练神经网络 情感分析 Keras Tensowflow			