

# 大数据与数据科学实战

## 技术结合数据 企业增长新引擎



加州大学伯克利分校哈斯商学院

**Berkeley** ExecEd | **Berkeley** Haas

# 概述



全球数字经济进入蓬勃发展阶段，大数据作为数字经济的核心内容和重要驱动力，也正以一种革命风暴的姿态闯入人们视野，并逐渐成为社会基础性战略资源，蕴藏着巨大潜力和能量。

随着大数据海量、多样化、快速化、价值化特征的逐步显现，加上数据资产逐渐成为现代商业社会的核心竞争力，众多企业已经意识到：掌握数据资产，进行智能化分析和管理的，是在企业竞争中脱颖而出的关键。麦肯锡公司曾进行过的一项研究表明，与非数据驱动的竞争对手相比，依靠客户分析的公司在获得新用户方面胜过竞争对手的可能性要高23倍，实现高于平均水平的获利能力的可能性要高19倍。于是在这个“数据密集型”的时代环境下，越来越多的企业开始重视和大数据相关的战略布局，并重新定义自己的核心竞争力，试图向数据驱动型企业转型。正如哈佛商业评论提到：大数据本质上是“一场管理革命”。

但是成长为“数据驱动型企业”的道路并不是一帆风顺的，它必遭遇巨大的挑战。New Vantage Partners的调查研究显示：

- 仅48%的企业认为他们在数据和分析方面有竞争优势；
- 仅31%的企业是数字驱动型企业；
- 建成数字文化的企业仅有28%。

实际上，除了需要企业领导层拥有先进的数据素养技能，阻碍实现数据价值和潜能的原因，不在于技术本身，而在于企业是否具备“将数据转换为可行性战略方案”的能力。

面对用户在网上源源不断产生得分散凌乱的数据，企业是否有能力通过这些数据洞察用户需求，挖掘其中的潜在信息就显得尤为重要。因此数据科学团队应运而生，他们通过大数据打通技术与数据之间的壁垒从而获取有价值的信息，推动业绩增长，并从长远角度上直接影响企业在市场竞争中的地位。

加州大学伯克利分校哈斯商学院推出的《大数据与数据科学实战》，在严格基于学科内容的基础上突出实践性，通过学习数据科学家日常工作中使用的工具、背后的统计学原理，以及众多数据科学应用案例和实操代码练习，最终帮助您自己或您的团队成员快速掌握数据科学思维与技能，使企业快速搭建数据团队并推动公司内以数据作为驱动的文化，提高公司整体竞争力。

66

“本课程能够帮助您真正提升数据素养。我们的重点不是数学，而是比谷歌分析报告更深入地挖掘数据。”

- Steve Tadelis (史蒂夫·塔德利斯)，哈斯商学院 James J. and Marianne B. Lowrey 商学讲座教授

## 适读人群

本课程在严格基于学科内容的基础上突出实践性,学员在亲身经历实践过程中学习先进的数据分析方法,无论您在任何行业或任何规模的公司或筹备创业,本课程均可帮您达成目标。

适读人群包含:

- 想要通过大数据快速建立数据科学团队,推动公司成长为数据驱动型企业的业务负责人
- 想要通过工具、原理以及练习快速入门数据科学领域,提升自身竞争力的职场人
- 想要将数据科学融入企业的人力资本、技术和管理挑战的企业负责人或管理层
- 需要对新入职数据科学团队的员工进行职能领域技能培训、管理数据团队的人力资源专家
- 需要为数据分析部门提供技术支持的业务人员
- 需要运用专业工具进行数据科学分析的经理们
- 技术导向型公司的企业团队
- 未来计划从事技术或IT项目工作的数据科学/大数据专业的学生

## 报名要求

- 需要具备数据分析基础或量化概念
- 需要具备数学基础知识 (统计学、函数、概率等)
- 了解基本数据统计工具

\*虽然本课程不需要您在上课前具备编程等专业知识,但为了您可以更好地理解 and 消化课程内容,需要您了解统计学、函数、概率等数学方面的基础量化概念。

# 课程介绍



加州大学伯克利分校哈斯商学院《大数据与数据科学实战》是一趟为期10周的学习之旅。由8个不同主题模块组成，旨在帮助各行各业在职专业人士或企业管理者深度学习和理解大数据科学和数据分析，并让学员在亲身经历实践过程中学习体验各种方法，以达到灵活运用数据的目标。这些对于任何希望释放大数据潜力的管理者或业务员工来说，都是至关重要的技能：

- 关于数据科学底层统计知识，本课程会在模块1-3中进行详细解析；
- 关于希望公司在未来依靠技术/人工智能转型为数据驱动型，但无数数据科学相关工具经验，本课程会在模块4-7中提供解答方案；
- 关于如何组建数据科学团队，本课程会在模块8中逐步说明。

完成本课程后，您将能更好地与数据科学团队、数据分析团队高效合作，帮助和指导数据科学家、数据分析师为企业更好地服务，实现更为可观的商业价值。

对于供职于大型企业组织的高级管理者，本课程可以帮助您：

- 快速入门数据科学领域，提升自身竞争力
- 掌握最常用的数据技能，从数据中获得洞见  
掌握数据素养技能，卓有成效地呈现和解释数据
- 快速提升专业度，获得伯克利哈斯商学院官方认证

对于正在领导团队的中小企业CEO、企业家，本课程可以帮助您：

- 快速组建数据科学团队，并具备评估依靠数据架构的战略技术决策的能力
- 将数据科学融入企业的人力资本、技术和管理挑战
- 加深公司对数据科学的理解，推动“数据驱动”的文化建立
- 获得伯克利哈斯商学院影响力，快速拓展商业人脉

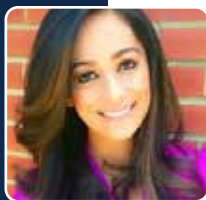


对于未来计划从事技术或IT业务的职场人/学生，本课程可以帮助您：

- 在未来快速适应数据科学相关工作
- 增强数据科学项目经验，提升个人竞争力
- 获得名校官方背书



# 学员感言



“本课程让我受益最深的是：使用Jupyter Notebook动手实践任务以及每周的教授坐班时间。”

- **Bitu Luliano**, 人才管理分析师



“我喜欢课程集视频课、动手实践任务和Jupyter Notebook应用练习于一体。”

- **Michael Wolff**, 执行总监, **Digital Platforms & Products**



“学习本课程后，我对数据科学有了深入了解，深度掌握了各种语言、模型、算法和值，并了解它们各自在实际应用中具备的优点和存在的问题。”

- **Saloni Sonawala**, 应用科学家



“我接触到好用的在线平台Jupyter Notebook，学习到现实世界的案例，课程还提供学习支持，这些都为精彩的学习之旅画上浓重的一笔。”

- **Dmitry Karablinov**, 业务发展经理



# 行业和企业实例 及应用研究

## 企业实例

由于加州大学伯克利分校哈斯商学院高管教育的教授,与硅谷及其周边地区众多具有行业代表性的全球知名科技公司联系密切,这些公司的科研及应用直接或间接地成为了本课程内容的来源:



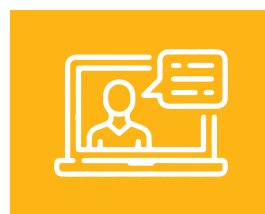
亚马逊



优步



易趣



盖洛普



StubHub

## 应用研究



### Rocket Fuel Ad Campaign (火箭燃料)

作为一家数字广告公司,火箭燃料使用大数据和人工智能来优化数字营销。使用Python运营广告宣传数据,使数据可视化,实施A/B测试,评判广告的成功。



### 首都共享单车

作为一家共享单车公司,首都共享单车致力于预测共享单车的单日使用总量。使用Python生成并分析摘要统计信息,创建回归模型,对共享单车的单日使用总量进行预测。



### IBM员工流失

使用IBM创建的虚拟数组对员工的离职率做预测。使用交互式平台Jupyter Notebook中的Scikit-Learn软件(一款免费的开源的Python机器学习软件),创建、训练并评估其它机器学习模型。

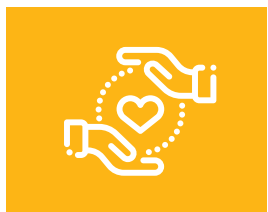
# 行业实例

体现在通过提升数据素养技能而受益的数据敏感性企业中。

典型行业包括：



金融科技/  
金融服务业



医疗保健业



信息技术业



制造业



零售业

注：本课程提到的企业名称或产品名称均为商标™或注册®商标名称。使用商标名称不代表和商标所有人之间有任何附属关系或赞助关系。



# 课程大纲

模块1：

## 概率决策

介绍数据科学和数据分析的基础概念，然后探索数据的基本知识。

- 区分分类数据和数值型数据
- 探索数据揭示信息的基本方式
- 类别变量和数值变量的相关性分析
- 医疗保健业实例：大数据管理式医疗 (HMO) 模式
- Jupyter Notebook, Python和Panda简介

模块2：

## 创建样本数据

了解数据分析报告常用的专业术语，学习样本分析的方法；借助样本分析，检验一类错误和二类错误和极限控制来分析企业决策的利弊。

- 了解样本类型、样本差异、样本质量
- 了解样本数据的基本概念及分类
- 识别并缩小样本数据中的偏差
- 实例展示：联合概率、边际概率、条件概率：Comcast、谷歌、Nextag

模块3：

## 假设检验

基于数据制定企业决策还依赖于表达清晰的假设，只有这样的假设才有助于统计检验。本模块会介绍假设检验的基础知识，包括统计比较、置信区间和误差范围。

- 实验设计的基本原则
- 识别并区分数据的单、双侧检验
- 使用4M模型解决问题 (动机、方法、机制、消息)
- 实例：24 Hour Fitness测试新专用饮食方案—测试对照组和治疗组

模块4：

## 利用样本数据推断

学习统计学中较为常见的直线模型和曲线模型。学习拟合线性模型的操作方法及适用范围：市场需求、定价、弹性。

- 了解使用直线模型和曲线模型的条件，学习解读直线模型和曲线模型
- 研究曲线 (非线性) 模型在车辆重量和燃油效率方面的应用
- 使用4M模型解决问题：美国费城的信用卡问题、犯罪问题、房价问题



---

模块5:

## 简单回归模型

简单回归模型是数据驱动型企业制定决策的核心模型。本模块的重点是了解简单回归模型的使用方法、模型有效的假设条件,以及如何使用模型帮助企业做出最佳决策。

- 了解简单回归模型的定义和应用;明确简单回归模型的适用范围
- 应用预测间隔并解释
- 确定影响回归模型的三大问题:数据变化、离群值和观测值之间的依存关系
- 零售业实例:使用回归模型解决连锁店的选址问题

---

模块7:

## 预测机器学习

学习基本概念,研究不同的应用,解密机器学习。

- 辨识机器学习的三种学习方法:有监督学习、半监督学习、无监督学习
- 研究机器学习的方法,包括有监督学习使用的“词袋模型”法
- 使用时间序列回归预测
- 网络安全实例:机器学习检测垃圾邮件

---

模块6:

## 多元回归模型

在基本概念的基础上进一步深入,定义多元回归模型,探索该模型的适用范围。

- 区分边坡斜率和偏斜率
- 利用多元回归模型分析:推断
- 总结多元回归模型的建模步骤
- 金融业实例:建立多元回归模型,解释索尼公司的股票投入产出
- 人力资源行业实例:使用修正回归模型分析工资数据,判断是否有同工不同酬的现象

---

模块8:

## 建立高效的数据科学团队

掌握基本概念和常用工具后,进一步深挖高效数据科学家团队所需要具备的能力,学习建设企业数字文化。本模块会指出需要避免的常见误区,并提供策略,帮助企业与数据科学家高效沟通。

- 学习建立高效的数据科学家团队的必要条件
- 继续探索数字文化建设
- 广告业实例:火箭燃料:转化率、优势、投入产出比、机会成本、A/B测试

\*注:为了帮助学员通过亲身体验实践学习数据技术,学会用数据制定更好的决策,课程设置了2次各为期1周的实验室学习课程,学员将有机会深挖数据。该实践活动是10周课程总体设计的一部分。

# 授课导师

## 史蒂夫·塔德利斯教授

哈斯商学院James J. and Marianne B. Lowrey

商学讲座教授



史蒂夫·塔德利斯教授在斯坦福大学任助理教授八年，后来到加州大学伯克利分校哈斯商学院任教。

曾在eBay研发实验室任总监、杰出经济学家（2011–13）。在亚马逊担任经济学与市场设计副总裁期间（2016–17），他与技术人员、机器学习科学家、业务领导携手合作，实现了将经济研究工具应用于多款产品和业务。至今仍是亚马逊经济学家团队一员，是亚马逊的兼职顾问。

史蒂夫·塔德利斯教授目前的研究领域：电子商务、组织经济学、采购合同、企业理论和产业组织理论、合同理论、博弈论。

### 部分荣誉和成就：

- 荣誉称号——全日制工商管理硕士课程Cheit教学奖, 2010–11
- 豪蒂亚斯奖——《比较经济学》期刊最佳文章, 2010–11
- 加州大学伯克利分校哈斯商学院Barbara and Gerson Bakar教职研究员, 2008–15
- 斯坦福大学Phi Beta Kappa 本科教学奖, 2005
- 斯坦福大学经济系优秀导师奖, 2002

# 沙查·卡里夫教授

哈斯商学院Benjamin N. Ward经济学教授



沙查·卡里夫教授是实验社会科学实验室 (Xlab) 的前系主任兼教务主任。其对行为经济学和实验经济学的研究, 为理解个人偏好、风险态度、时间态度等问题提供了新工具, 为决策提供几近全面的信息。

学术方面, 教授曾在斯坦福大学、普林斯顿大学、牛津大学、剑桥大学、欧洲大学学院、挪威经济学院等学府担任客座教授。教授还是Capital Preferences公司的联合创始人、首席科学家, 这家公司变革了企业投资建议、贷款、保险、人才市场解决方案。

## 部分荣誉和成就:

- 加州大学伯克利分校哈斯商学院 Earl F. Cheit 教学杰出奖, 2012
- 获美国斯隆基金会研究资助, 2009
- 加州大学伯克利分校社会科学系杰出教学贡献奖, 2006-07
- 加州大学伯克利分校经济系研究生会优秀导师奖, 2006-07

# 认证证书

顺利完成本课程的学习后,学员将被授予由加州大学伯克利分校哈斯商学院高管教育认证的证书。  
本课程成绩分合格和不合格,学员分数必须达到80分才能被判定为合格并获得结业证书。



注:成功学完本课程后,我们将按照您报名时登记的姓名签发电子版认证证书,并使用电子邮件发送到您的邮箱。证书中使用的图像仅用于说明目的,加州大学伯克利分校哈斯商学院高管教育可酌情更改。



BerkeleyHaas

## 关于加州大学伯克利分校哈斯商学院

哈斯商学院隶属于加州大学伯克利分校,也被称为伯克利哈斯商学院,是享誉全球的著名商学院,同时也被誉为全球入学难度最大的知名商学院之一。该校地处硅谷,而硅谷作为众多创新型知名企业的集聚地和新技术的发源地,给伯克利哈斯商学院提供了源源不断的理论依据和行业案例。作为众多全球先进企业数字化战略层面的重点合作对象,伯克利哈斯商学院也以其丰富的技术和战略经验助力大量企业完成一系列数字化相关的战略计划。

加州大学伯克利分校成立于1868年,是世界顶尖公立大学。它不仅是加州大学10个分校中校史最长的一个,而且其教学质量、科研成就、师资、硬件设备和学生质量也是10个分校中最顶尖的,在全世界也一直名列前茅。由于地处硅谷边缘,学校毕业生中出现了许多的新型高科技富翁,例如英特尔公司创始人戈登·摩尔、安迪·葛洛夫都是加州大学伯克利分校的校友。在每年的全美公立大学排名中,伯克利校区长期独占鳌头,位居第一。该校的本科生获得博士学位的人数远高于其他任何美国的大学。





## 课程安排

10周学习时长(线上学习8周, 实验学习2周)  
每周6-8小时

\*学习模块、完成作业和参加实时学习支持会议所需的时间长度取决于学员的技术能力和背景。此为预估顺利学完课程平均需要花费的时间。

## 关于Eruditus Group

Eruditus Group (<https://emeritus.org/>) 成立于2010年, 专注于和世界著名的商学院合作, Eruditus Group 旗下品牌“ERUDITUS”和“EMERITUS”, 前者提供线上线下混合培训课程, 后者专注于线上培训课程。自成立以来, 通过与全球知名院校开展创新性地通力合作, 实现了校内、校外和在线授课相结合的方式, 为全球高管提供工商管理及领导力培训等课程。

伯克利哈斯商学院高管教育与EMERITUS合作。EMERITUS通过线上的录播和直播教学方式以及实操项目, 为每年超过160个国家和地区的66000多名学员提供高质量的管理培训。“优仕研习社”为“EMERITUS”旗下的中国品牌, 致力于让更广泛的学员接触到伯克利哈斯商学院的权威管理课程。





## 联系课程顾问

电子邮件: [berkeley@emeritus.org](mailto:berkeley@emeritus.org)

电话: 400 1008 874 (中国)

我们会尽可能在24小时之内答复。

如果您在周末和节假日期间咨询, 我们会在72小时内答复。

\*“ERUDITUS”、“EMERITUS”和“优仕研习社”均为 Eruditus Group 旗下品牌。“优仕研习社”的主体为上海择藤教育科技有限公司, 在法律允许的范围内择藤享有以上资料的解释权。

**Berkeley** Executive  
Education

**BerkeleyHaas**

**E** EMERITUS