

大数据背景下

高校教学与管理

目录

Content

1 大数据时代的学、教、管

2 国内外大数据在高等教育的应用案例

3 在日常教育教学与管理中应用大数据



大数据时代的学、教、管

➤ 什么是**大**数据？

大 数 据

Big Data

➤ 认识大数据时代



什么是大数据

➤ 认识大数据时代

用大数据追女神

优酷

➤ 认识大数据时代

大数据帮助奥巴马成功实现连任

奥巴马的数据团队对数以千万计的选民邮件进行了**大数据挖掘**，精确预测出了更可能拥护奥巴马的选民类型，并进行了有针对性的宣传，从而帮助奥巴马成为了美国历史上唯一一位在竞选经费处于劣势下实现连任的总统。



➤ 认识大数据时代

在某一个周末……



7: 00, 你被手机闹钟叫醒。昨晚你带着一款小型可穿戴设备睡觉的。这个设备连接着你手机里的一款**大数据的APP**, 你打开它就可以看到你昨晚睡觉时翻身次数、心跳和血压状况。根据测量结果, 它建议你今天出门之前多喝点橙汁类的饮品来补充维生素。



9: 00, 今天你要带朋友到汉口的步行街逛逛, 你打开某互联网公司的大数据产品“**XX预测**”, 看看步行街今天预计会有多少人, 再看看武汉今天的交通预测。“**XX预测**”根据以往用户定位请求信息建议你乘2号线地铁去步行街。



12: 00, 逛了一圈, 你和朋友都累了, 想找个地方吃饭。你打开**大数据软件**, 寻找附近的餐馆。通过该软件, 你可以提前看到餐馆的视频环境, 看看是否人多。大数据还可以把你脸的部分打成马赛克, 你不用担心个人信息泄露。

➤ 认识大数据时代



14:00, 吃过午饭, 你想去附近的公园玩玩, 但你不知道应该去解放路公园还是去中山公园。你又打开“**XX预测**”, 希望它帮你分析一下, 哪个公园相对不太拥挤。根据结果, 你去了中山公园。



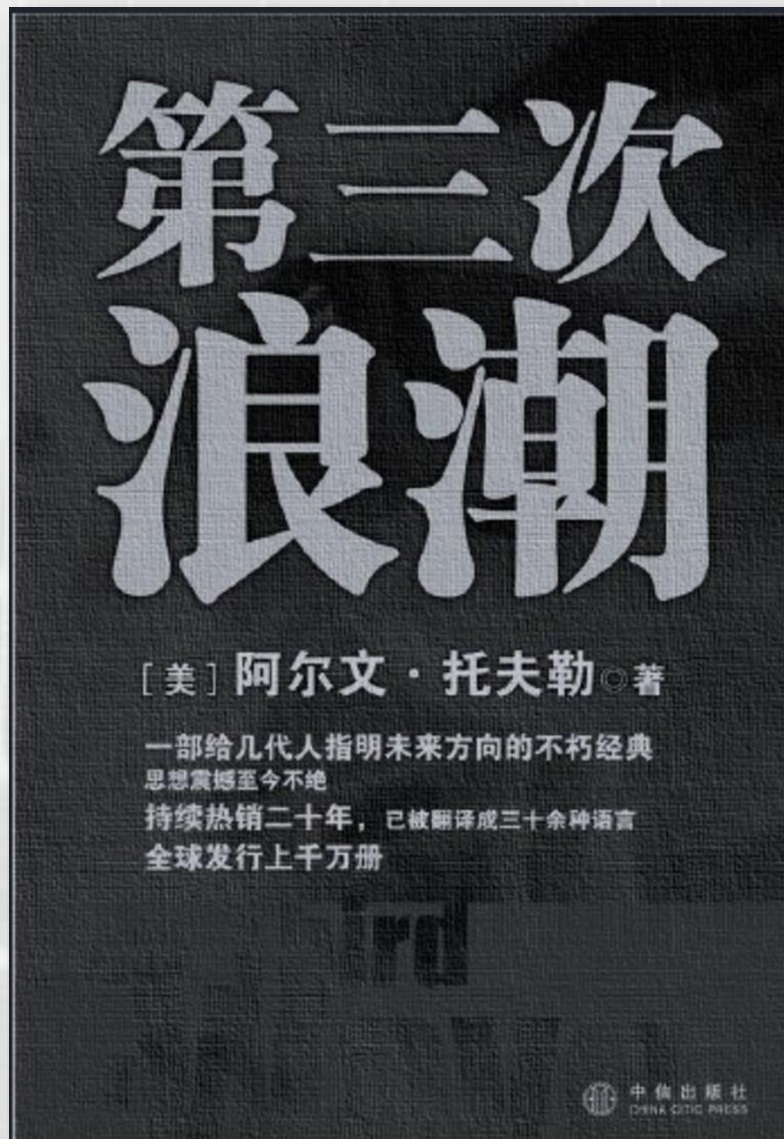
16:00, 你正在公园里休息, 收到了催缴电话费的短信。你很好奇自己过去三年每个月的消费记录。但过去运营商只能让你查到六个月以内的消费信息。因为中国移动至少有5亿用户, 每天至少能产生10次计费记录。每天50亿的计费记录, 一年下来就是1800亿条记录。这是一个**大数据的存储和归纳技术**难题。但在大数据时代, 过去几年的电话通讯信息都可以查到了。

18:00, 你回到了家, 你的可穿戴设备告诉你, 今天你在室内和室外的时间分别都是多少, 你一天内吸入了多少雾霾。

22:00, 晚上睡觉的时候, 你家的孩子哭闹起来。你把孩子的哭声录入一个大数据库软件中。软件能告诉你孩子为什么哭。是饿了, 还是哪里不舒服, 还是说只是想撒撒娇……



➤ 什么是**大数据**？



早在1980年，未来学家托夫勒在其所著的《第三次浪潮》中热情地将“大数据”称赞为“第三次浪潮的华彩乐章”。

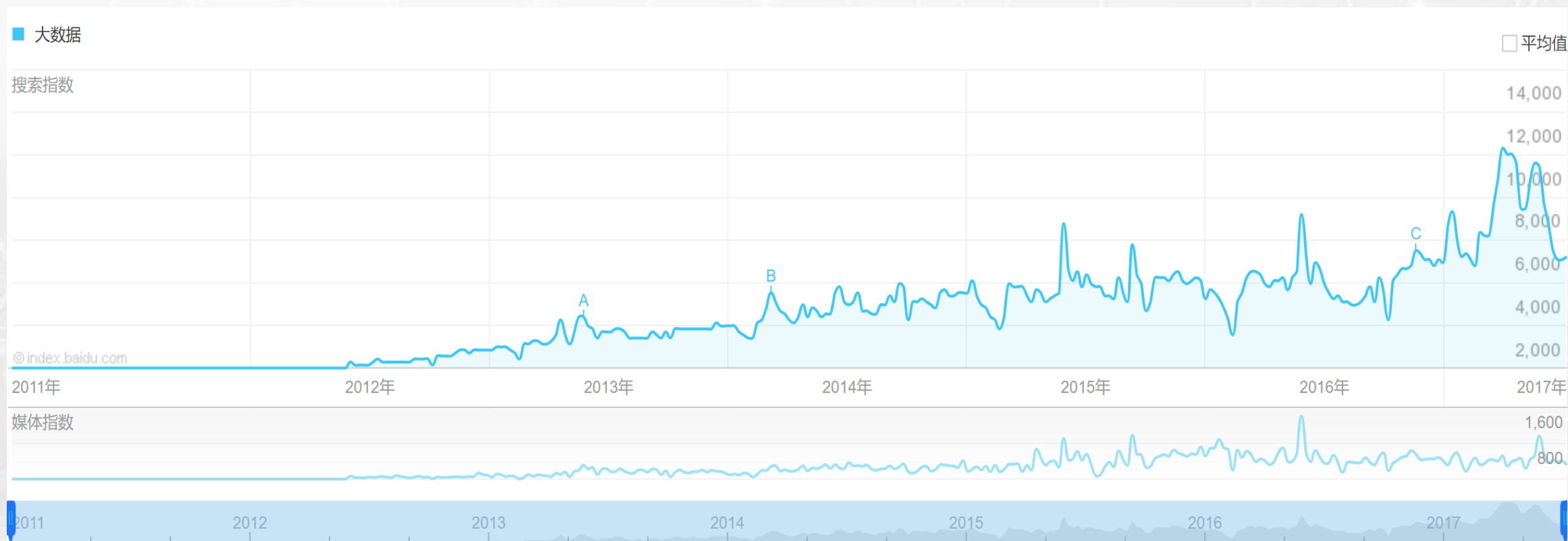
➤ 什么是大数据？

2008年9月，《自然》杂志推出了名为“大数据”的封面专栏。



➤ 什么是大数据？

2009年开始，“大数据”逐渐成为互联网技术行业中的热门词汇。随后，人们对“大数据”的关注度猛烈增长。



2011年至今在百度中以“大数据”为关键词的检索次数

➤ 什么是**大数据**？

一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合，具有**海量的数据规模、快速的数据流转、多样的数据类型和价值密度低**四大特征。

——麦肯锡全球研究所

4V:

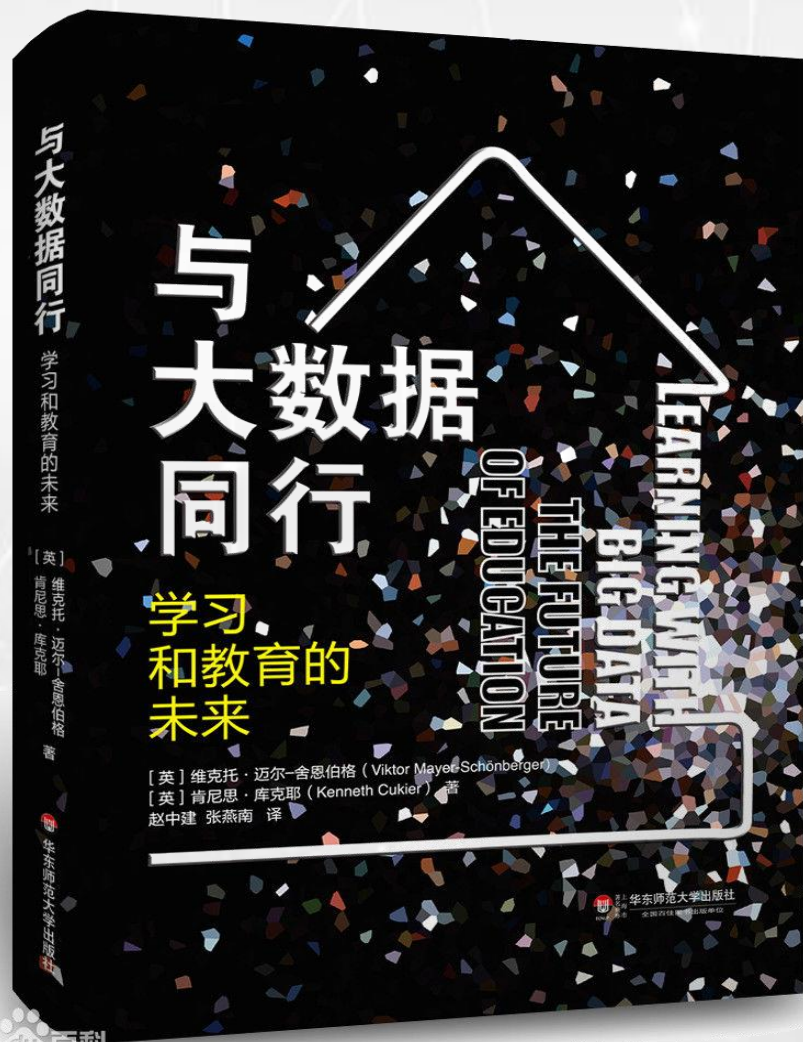
Volume: 数据体量巨大。

Variety: 数据类型繁多。包括数字、网络日志、音频、视频、图片、地理位置信息等。

Value: 价值密度低。需要通过强大的机器算法更迅速地完成数据的价值“提纯”。

Velocity: 处理速度快。

➤ 什么是**教育**大数据？



对现代教育产生的大数据进行分析，深入挖掘信息规律，反馈教学活动及其管理，科学高效地实现**教学精准化、学习个性化、环境智能化、管理智慧化**的总体目标。

➤ 教育大数据能做什么？



表 1-1: 学习效果

	传统在线系统	自适应系统
前测成绩 (标准差)		
后测成绩 (标准差)		
提分 (标准差)		
提分差别 (95% 置信)		
p 值		

表 2-1: 学习效率

	传统在线系统 (N=111)	自适应系统 (N=109)
总提分	23.0	569.0
总学习时间 (小时)	213.2	188.6
学习效率 (总提分 / 总学习时间)	0.11	3.02
学习效率倍数 (自适应系统 vs. 传统在线系统)		27.97
p 值		<0.001

教学精准化与个性化

➤ 什么是**教育**大数据？

教学环境智能化

➤ 什么是**教育**大数据？

学校管理智慧化



国内外大数据在高等教育的应用案例

➤ 教育大数据与学习分析



单纯的成绩单不过是一个**数字**

数据



+ 态度 **+** 家 **+**

➤ 教育大数据与学习分析



A同学

做对第4题



跳过第7、9题

做对第4题的同学几乎不可能做错第7题与第9题



B同学

做错第5题



继续做第6、8题

数据显示，做错第5题的人很可能会做错后两题

➤ 教育大数据与学习分析

“学习分析技术是通过测量、收集、分析和报告有关环境的数据，用以理解和优化学习、学生及其学习环境的技术。”

——2011《学习分析和知识国际会议》

“学习分析”这一术语最早来源于商业领域，通过对商业活动进行系列数据分析，商家可以对消费者趋势和消费行为进行预测。随着互联网的发展大数据的出现，学习分析逐渐被运用于大规模的教育教学领域。探索大数据的科学分析，以改善教师的教学方法、提升学生的学习积极性及学习质量等，进而帮助学校、教师有效调整课程、教学和评价系统。

——《国际教育信息化发展报告（2013-2014）》

2011年地平线报告中学习分析被列为未来年的主流技术，2012、2013年地平线报告中学习分析技术均被列为未来2~3年的技术，并在实践中开始推广了应用。

➤ 教育大数据与学习分析

大数据在教育领域中的应用，主要指的是在线决策、学习分析、数据挖掘三大要素，其主要作用是进行预测分析、行为分析、学业分析等的应用和研究，大数据含义指的是对学生学习过程中产生的大量数据（数据来源包括两方面，即显性行为 and 隐性行为，其中隐性行为包括论坛发帖、课外活动、在线社交等不直接作为教育评价的活动，显性行为包括考试成绩、作业完成状况以及课堂表现等）进行分析，大数据模型以及显示的数据能够为学校和教师的教学提供参考，及时、准确的评估学生的学业状况，发现学生潜在存在的问题，进而预测学生未来可能的表现。

1) 构建学习者经验模型

- 收集学习者的学习满意调查问卷
- 课程学习中的存留数据、表现、行为、选择等，构建学习者的经验模型
- 构建学习者体验模型，进行线上课程评估

2017/7/17

2) 建立学习者行为模型

- 收集学习者在学校情景中学习行为变化的情况
- 学习者完成课程学习的状况
- 学习者在网络系统中花费的学习时间以及学习者的考试成绩等数据

3) 构建学习者知识模型

- 学习者请求帮助的性质和数量
- 学习者回答问题花费的实践
- 学习者回答错误的重复率
- 学习者回答的正确率等
- 通过数据挖掘和学习分析构建的学习者知识模型

4) 构建领域知识模型

- 对现有的领域知识进行重新建模
- 研究学习者与知识点、学习单元、课程等学习内容之间的关系
- 采集和处理学习者的相关数据，画出学习者的学习曲线

5) 构建学习者档案

- 收集学习者的基本学习信息，建立基本信息数据
- 通过数据挖掘、学习分析和机器学习算法，根据学习者的学习特征，将相同学习特征的学习者进行分组和聚类

6) 教学策略分析

- 收集的学习者的信息进行分析，探索学习系统中各种组件的功能
- 分析学习者学习结果与教学策略之间的关系，然后对教学策略进行分析和总结

7) 其他应用

- 大数据在教育领域的运用还包括个性化学习、自适应学习系统、以及趋势分析等方面
- 通过大数据的收集、分析与处理，为学习者提供个性化学习、与自适应学习的环境

1. 从学生方面来说，学习分析技术在了解学生学习现状之后，通过分析学生数据，找出相关问题，对学生学习过程进行优化，帮助学生培养良好学习习惯，从而达到学生自我学习的目的。
2. 从教师以及管理人员方面来说，学习分析技术可以评估教学课程和相关机构，帮助同步改善学校既定考核方式，深入分析教学数据，为教师帮助学生解决实际问题指明教学不足和更优方法。
3. 从研究人员方面来说，学习分析技术是一种研究学生和网络学习的有效工具。
4. 从技术开发人员方面来说，学习分析技术管理系统各模块各不相同的使用频次和路径能有效指导系统界面的相关优化设计，并可以完善系统日志相关管理功能。

(1) 内容分析法

- 是一种对文献内容作客观系统的定量分析的专门方法
- 是弄清或测验文献中本质性的事实和趋势
- 建立有意义的类目分解交流内容

(2) 话语分析

- 利用话语的生产模式来行使权力
- 是权力的替代品
- 掌握话语权的人掌握社会规范

(3) 社会网络分析

- 来分析其他类型数据对关系数据的影响
- 数据主要分为两类，属性数据和关系数据

(4) 聚类

- 是在“物以类聚”这一理念之上
- 专门
- 针对大量的样品，按各自的特性进行相对合理的分类

(5) 预测

- 根据已有数据或信息对整体或个体数据发展态势做出合理预测

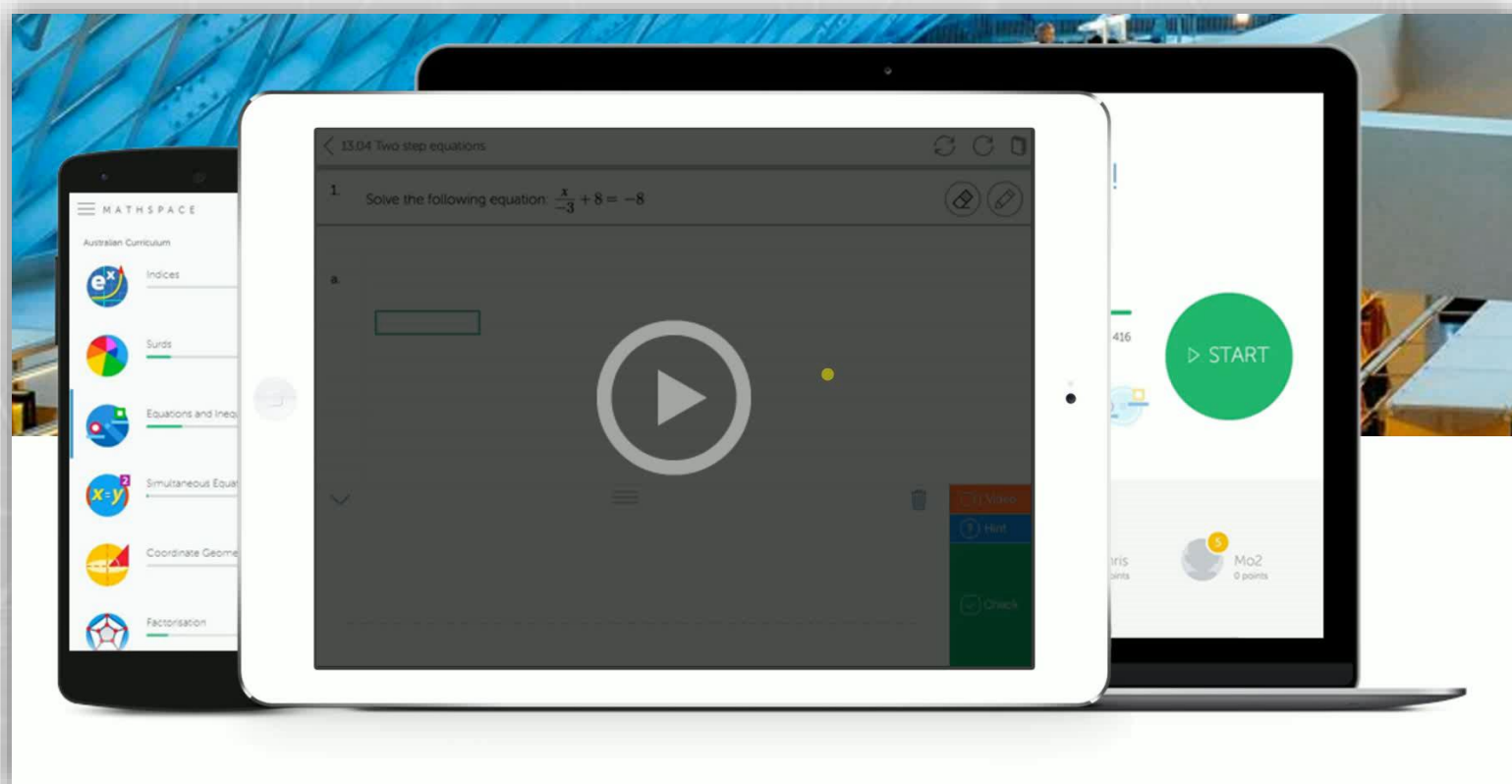
(6) 关系挖掘

- 在信息领域当中发展最快的技术
- 数据的高维特征并不是传统数据分析方法可
- 以轻易解决的

(7) 文本挖掘

- 文本数据中抽取有价值的信息和知识的计算机处理技术
- 文本数据挖掘是数据挖掘的一个分支

➤ 教育大数据与学习分析



Mathspace 平台监控学生对数学问题的思考 并提供个性化的反馈及为教师提供分析报告。

官网链接：<https://mathspace.co/>

➤ 教育大数据与学习分析

United States ▾ Contact Us Bookbag

Pearson PreK-12 Education Higher Education Industry & Professional About Us Search

Subject Catalog ▾ Products & Services for Teaching ▾ Products & Services for Institutions ▾ Customers ▾ Events Why Choose Pearson? ▾

Higher Education > Products & Services for Teaching > Learning & Engagement Tools > Learning Catalytics SIGN IN REGISTER

Learning Catalytics

Engage your students in team-based and group-learning activities with this interactive student response tool.

Get training and support >

Learning Catalytics

Features

User Stories ▾

Foster student engagement and collaborative learning

Learning Catalytics is an interactive student response tool that encourages team-based learning by using students' smartphones, tablets, or laptops to engage them in interactive tasks and thinking.

Learning Catalytics (LC) 由2位哈佛教授和一位博士后在2011年开发的一种基于云技术的学习分析评估系统，2013年被培生公司收购，现逐步推广至美国教师的课堂教学之中。

LC支持教师对学生学习结果的监控以提供实时反馈和支持学生的自动分组。

官网链接：<https://learningcatalytics.com/pages/stories>

➤ 教育大数据与学习分析

Knewton 平台自适应学习

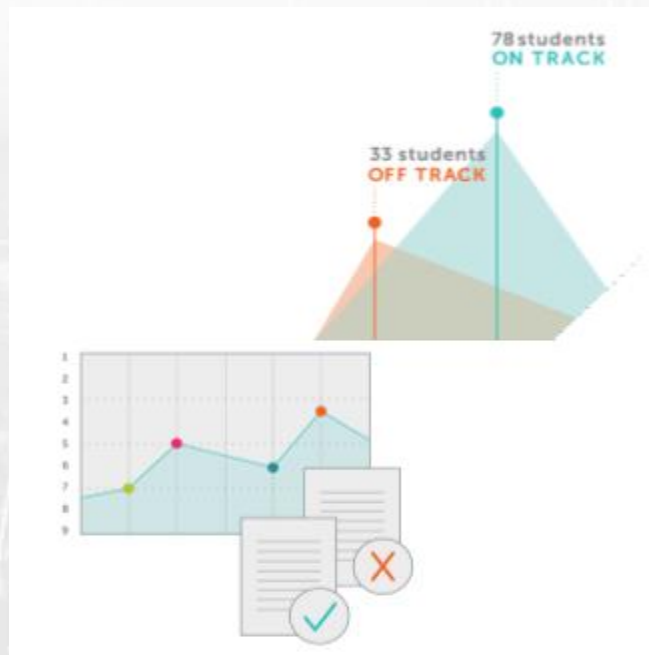
The Knewton API

Recommendations



个性化推荐

Analytics



学习结果分析

Learning histories



学习过程记录 (学习档案)

➤ 教育大数据与管理

普渡大学课程信号项目

目的：

- 1.应对普渡大学内日益下降的新生保有率
- 2.应对日益延长的毕业生毕业周期的危机

过程：

通过数据挖掘和统计预测模型，根据多个变量来预测学生是否能够完成课程。在课程进行的过程中，以交通信号指示灯的方式，让学生了解学习状况。同时教师可以提供给学生有效的反馈信息，引导学生使用合适的资源等来提高成绩。

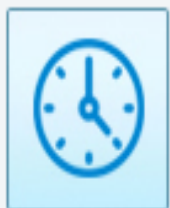
The screenshot shows the 'Signals' web application interface. At the top, there is a green header with the 'Signals' logo and navigation links for 'Home', 'About', 'Help', and 'Logout'. Below the header, the user's name 'Mary Major' is displayed, along with three buttons: 'Detailed Report', 'Effort Tracker', and 'Help Resources'. The current semester is 'Fall Semester'. A table lists courses with their progress indicators (green, yellow, or red dots) for three intervals (Int 1, Int 2, Int 3).

Course	Int 1	Int 2	Int 3
BIOL 101	●	●	●
GS 101	●	●	●
SPAN 310	●	●	●
STAT 303	●	●	●
COM 150	●	●	●

At the bottom of the page, the Purdue University logo is displayed, along with the text 'PURDUE UNIVERSITY' and a copyright notice: '© 2009 Purdue University. An equal access, equal opportunity university.'

➤ 教育大数据与管理

基于大数据的学生行为分析



晚归未归



电器收缴



消费异常



不在校



非正常上网



假期留校
区域分布



假期留校
预警



在线上网

门禁系统

一卡通

上网管理

请假信息

舆情监控

➤ 教育大数据与管理

● 以质量提升为导向，建立评估决策预警机制

实时监测
在线评教、满意度测评



能力判断	<p>传统学习模式</p> <p>市面常用的纸质试卷，对于低能力与高能力的学生而言，无法有效满足这些考生能力估计的需求。</p>	面向学科的能力评价	<p>自我调整测验，可依学生程度出题，即时测评，找出学习盲点。</p>
能力分析	<p>仅以考试分数判断能力高低，无法得知学习弱点。</p>		<p>重点能力诊断搭配精华处方，有效找出学习弱点。</p>
学习后验收诊断	<p>在相同学习范围重新出题，以分数高低判断学习是否进步。</p>		<p>以电脑自我调整出题方式判断学习成效，了解整体学习成效与问题。</p>

教学质量智能诊断和评价

➤ 教育大数据与管理

华中师范大学勤助中心：
“发现你上个月餐饮消费较
少，不知是否有经济困
难？”

华中师大利用预警系统跟踪学生的餐饮消费数据，发现低于警戒值就发出短信慰问，以确定学生是否有经济困难、是否需要帮助。



教育大数据与管理

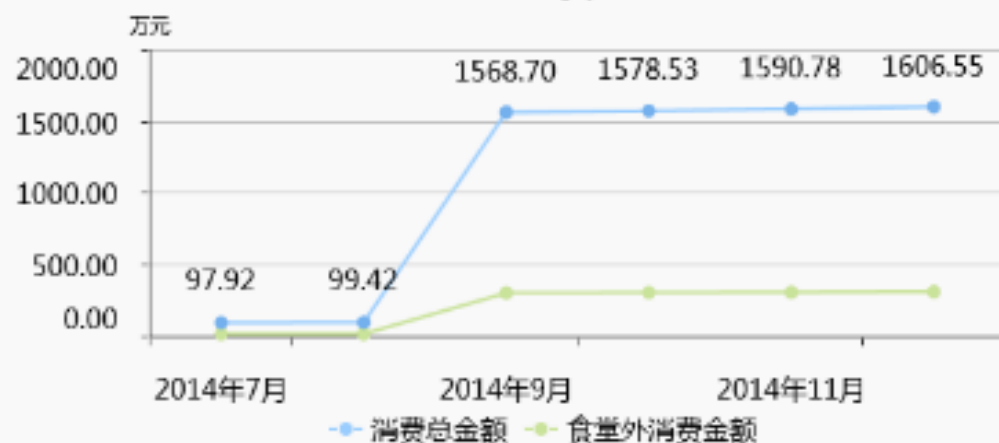
消费异动情况

片区	超变动率 上限人数	次变动率 下限人数	月均每天消费 金额大于50元 人数	月均每天消费 金额小于5元人 数
东一片区	352	19	12	8
东二片区	364	18	13	4
南湖片区	289	16	21	12
西区片区	159	8	10	3

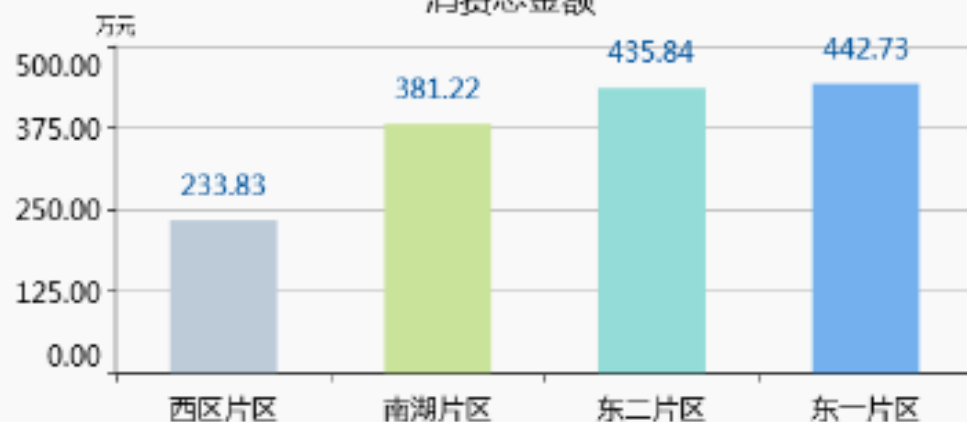
平均每顿消费金额



近半年消费折线



消费总金额



学生消费行为数据分析

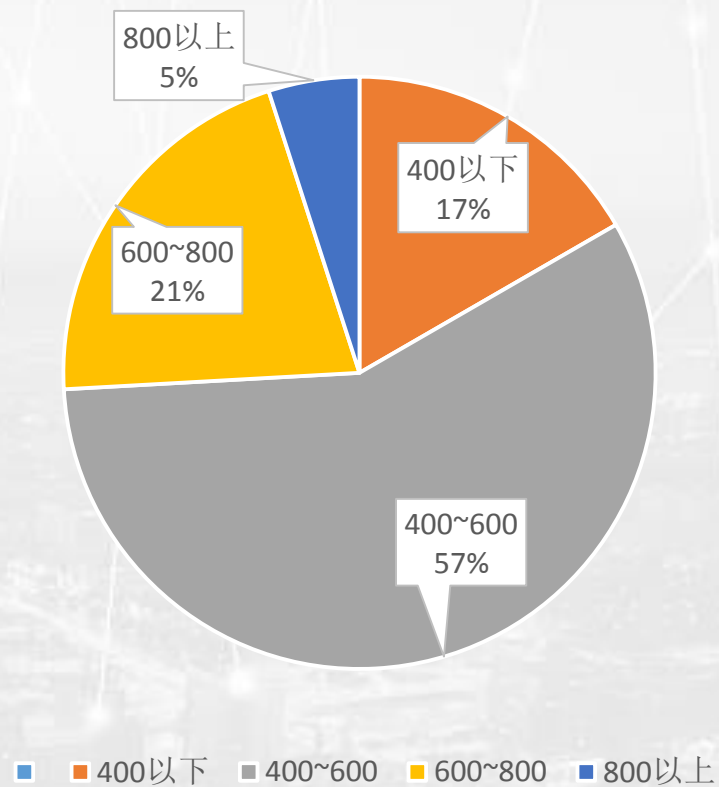
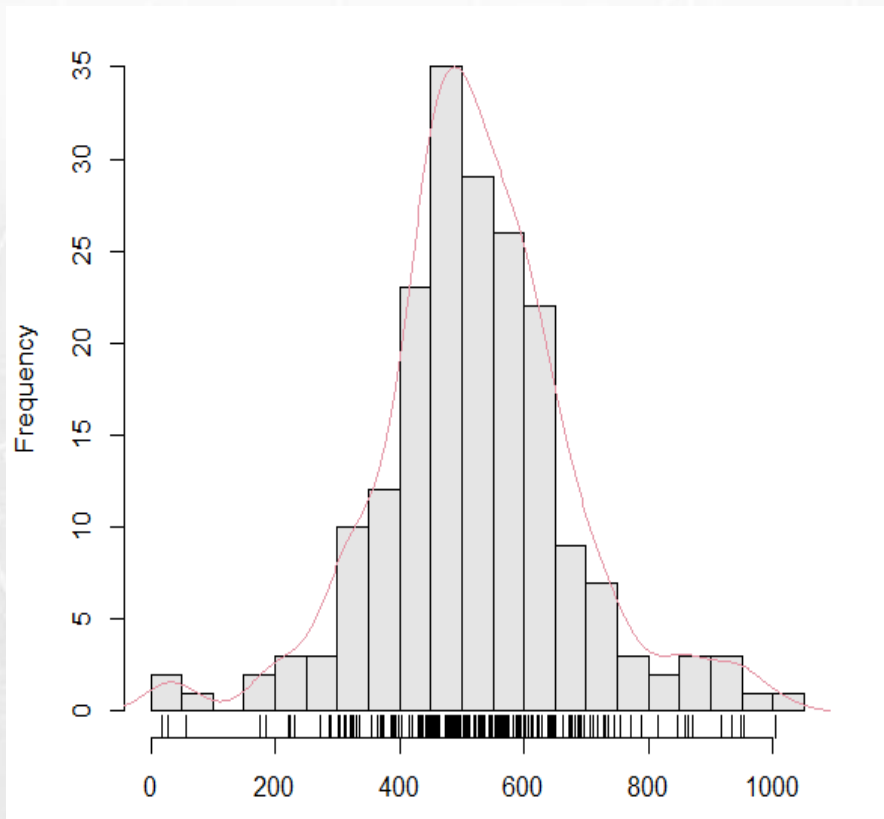


图1 月平均消费金额

➤ 教育大数据与管理



辽宁卫视

百秒妙评

高清

大连 晴 28℃ - 17℃

06:56

百秒妙评

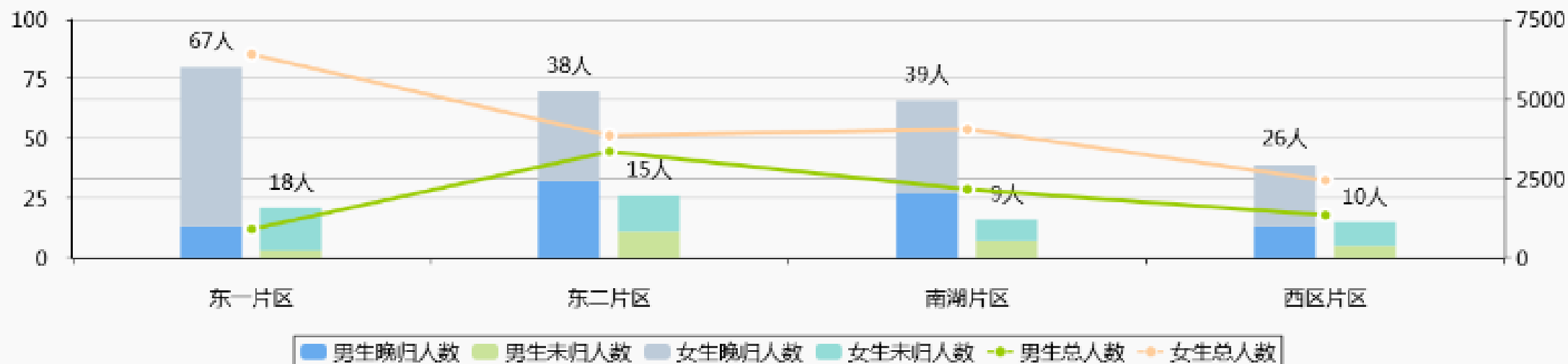
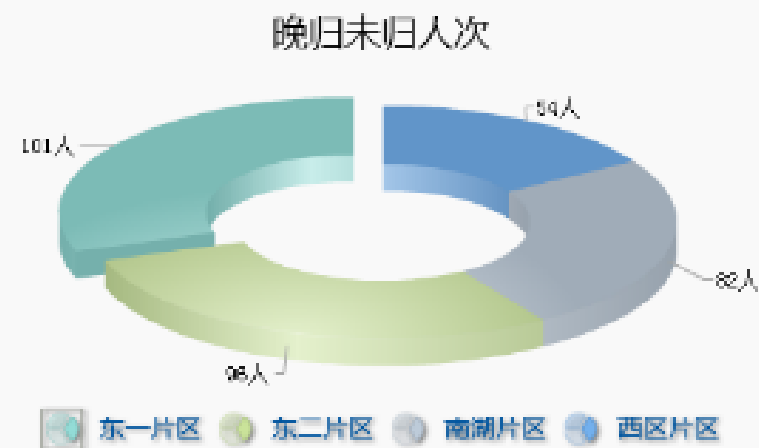
新西兰小城举办“人肉中国人”大赛 胜者免费游中国

文物平台，大家足不出户就能在手机等移动终端对文物藏品进行全方位、多角度的观赏

教育大数据与管理

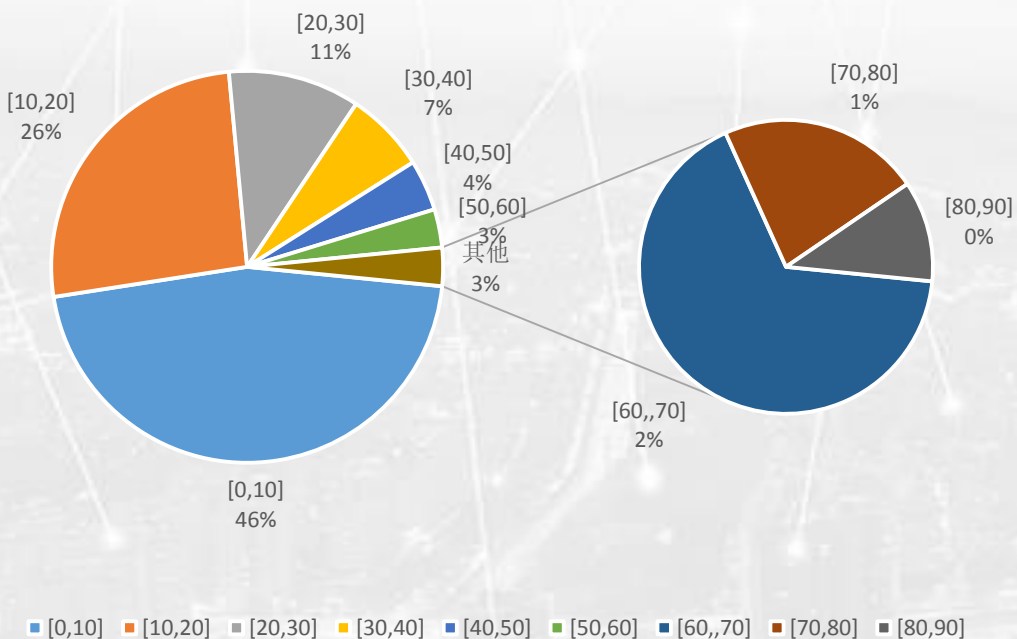
2015年01月22日各片区晚归、未归情况

片区	总人数	未归人次	晚归人次	总人次	男异常率	女异常率
东一片区	7300	21	80	101	1.78%	1.33%
东二片区	7205	26	70	96	1.29%	1.38%
南湖片区	6203	16	66	82	1.58%	1.19%
西区片区	3778	15	39	54	1.34%	1.48%



学生学习行为 数据分析

图书馆入馆次数



图书馆入馆次数分布

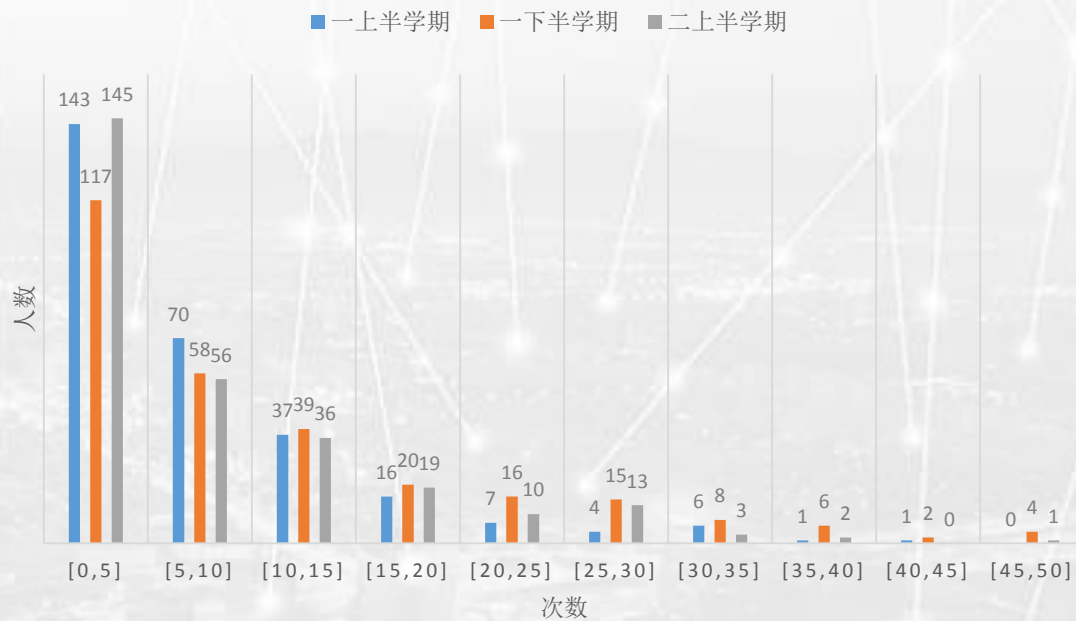
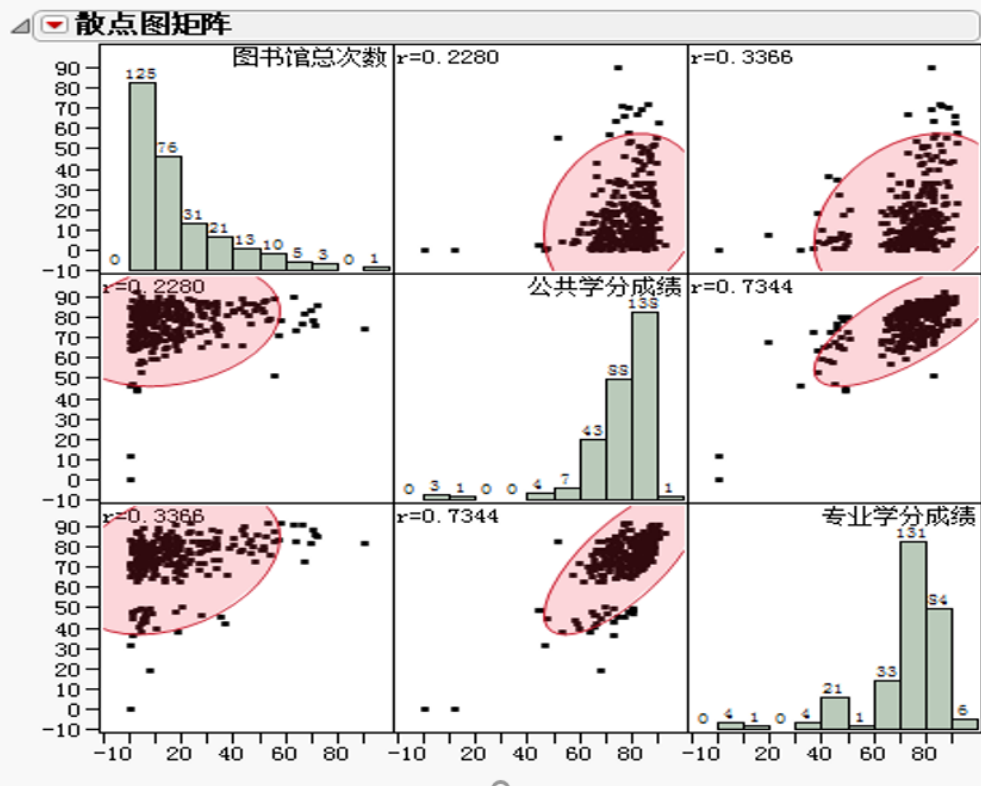


图2 图书馆入馆次数

教育大数据与管理

经常进入图书馆有利于提高学业成绩



多元

相关性

	图书馆总次数	公共学分成绩	专业学分成绩
图书馆总次数	1.0000	0.2280	0.3366
公共学分成绩	0.2280	1.0000	0.7344
专业学分成绩	0.3366	0.7344	1.0000

图 15 学生图书馆入馆次数与学业成绩相关性分析结果图

➤ 高等职业教育的特殊性

特殊的学生

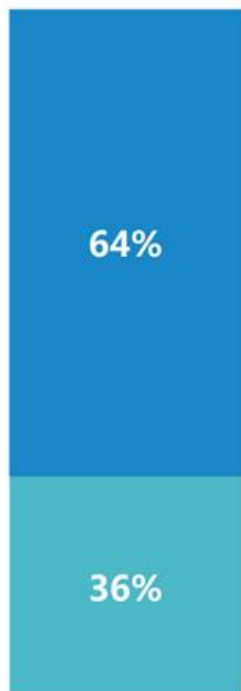
学科教育中不够成功的过来者

特殊的培养目标

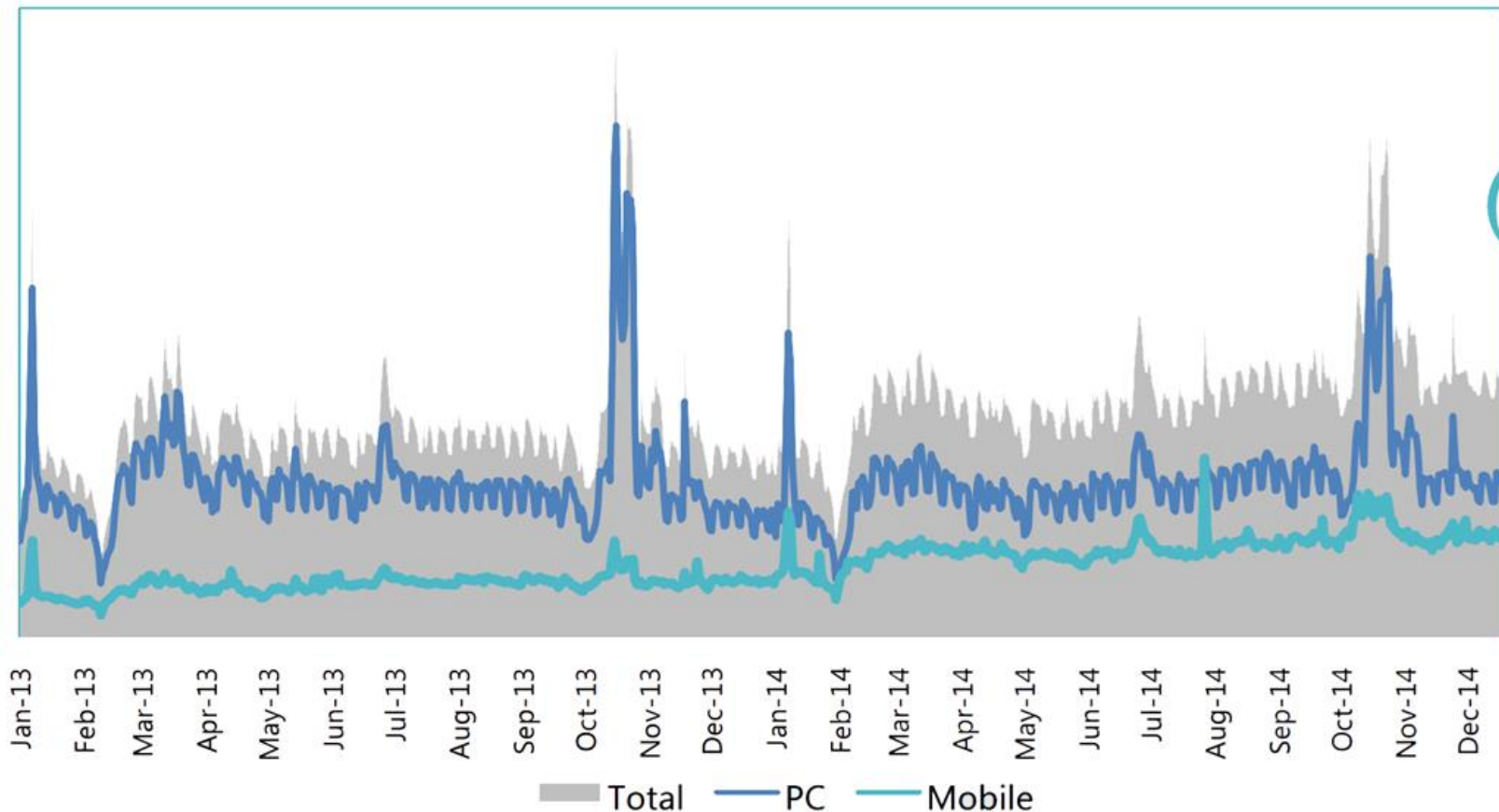
- ✓ 不太适应逻辑性强的学科性知识学习；
 - ✓ 对理论学习的主动性和积极性不高；
 - ✓ 更适应形象思维，适合“在做中学”；
 - ✓ 学习动机是谋求一份理想的工作；
 - ✓
-
- ✓ 有能力会“怎么做”，“怎么会做得更好”；
 - ✓ 知道必要的“是什么”，“为什么”；
 - ✓

互联网时代在线职业教育中的大数据

2014年
流量占比



职业教育行业搜索趋势(2013.1-2014.12)



01

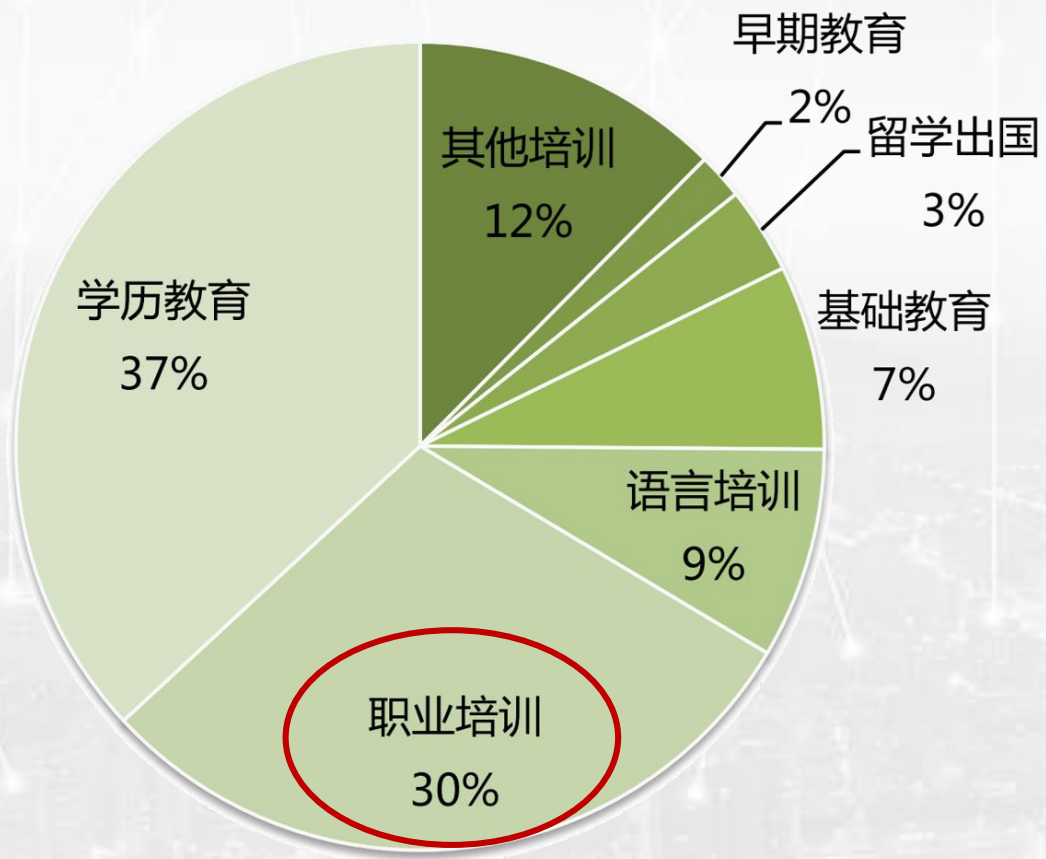
职业教育的年度搜索量增长15%，高于大盘，且移动端需求占比高于整体教育行业

02

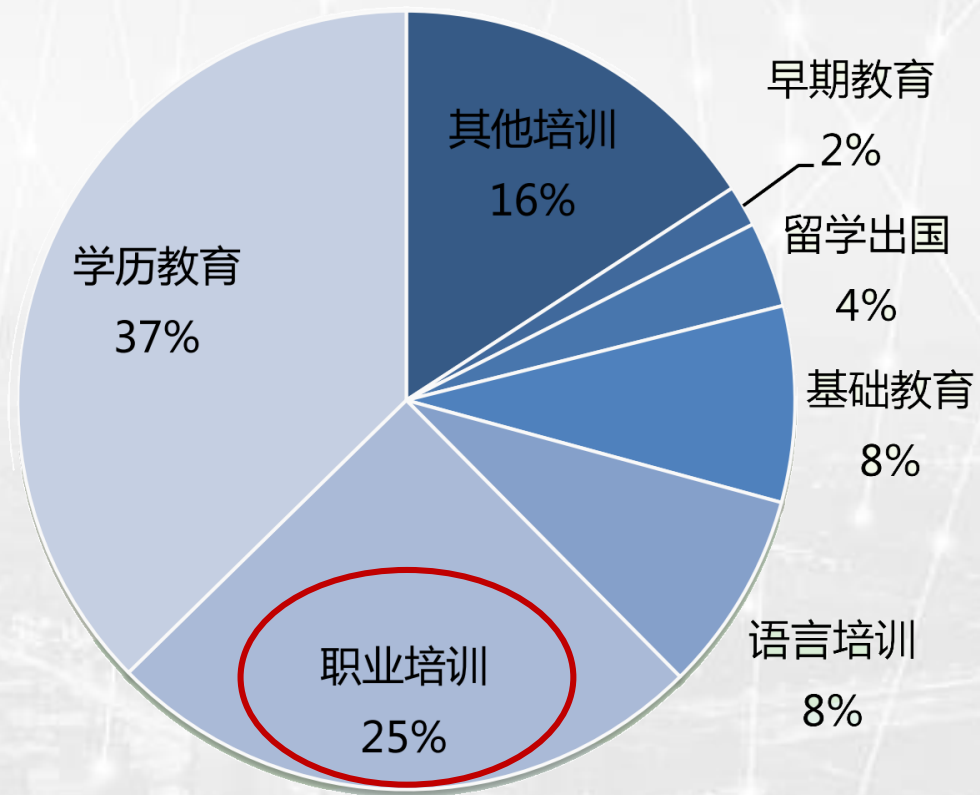
职业教育每年三个旺季，3月为春节后返城求职人员，寻求职业能力进一步提升；6月为暑期学生人群，主要为高考失利学生寻找职业培训的出路；10月为公务员国考笔试公告的发布，备考人员关注度暴增

互联网时代在线职业教育中的大数据

2014年PC端各子行业需求占比



2014年移动端子行业需求占比



数据来源：百度PC&移动端201301-2014112教育行业子行业流量数据

➤ 互联网时代在线职业教育中的大数据

工作日及周末日均搜索分布对比

搜索量时段分布

互联网给高等职业教育带来的挑战 尤其之大

周末
43%

工作日
57%

10点

15点

20点



在日常教育教学与管理中应用大数据

➤ 教育大数据的特点



面临的挑战：数据的时效性；数据如何收集；数据如何结构化以支持数据挖掘

➤ 积累与运用教育大数据，我们如何做？

大数据要能发挥其作用，**数据的积累**是基础，数据挖掘要能起到该有的作用，需要好的算法和规范合适的数据格式。

数据不是问题，应用才是关键

平时多思考：针对待解决的问题，我们需要什么样的数据？如何收集与保留这些数据？收集这些数据后如何分析？期待得出什么结论？

➤ 大数据在高等职业教育中可以这么用

差异化教学

教育/教学不该是孤立的，不能是只凭经验的，教育/教学决策需要大数据作为证据。

.....



➤ 大数据在高等职业教育中可以这么用

差异化教学



传统的教室



校园



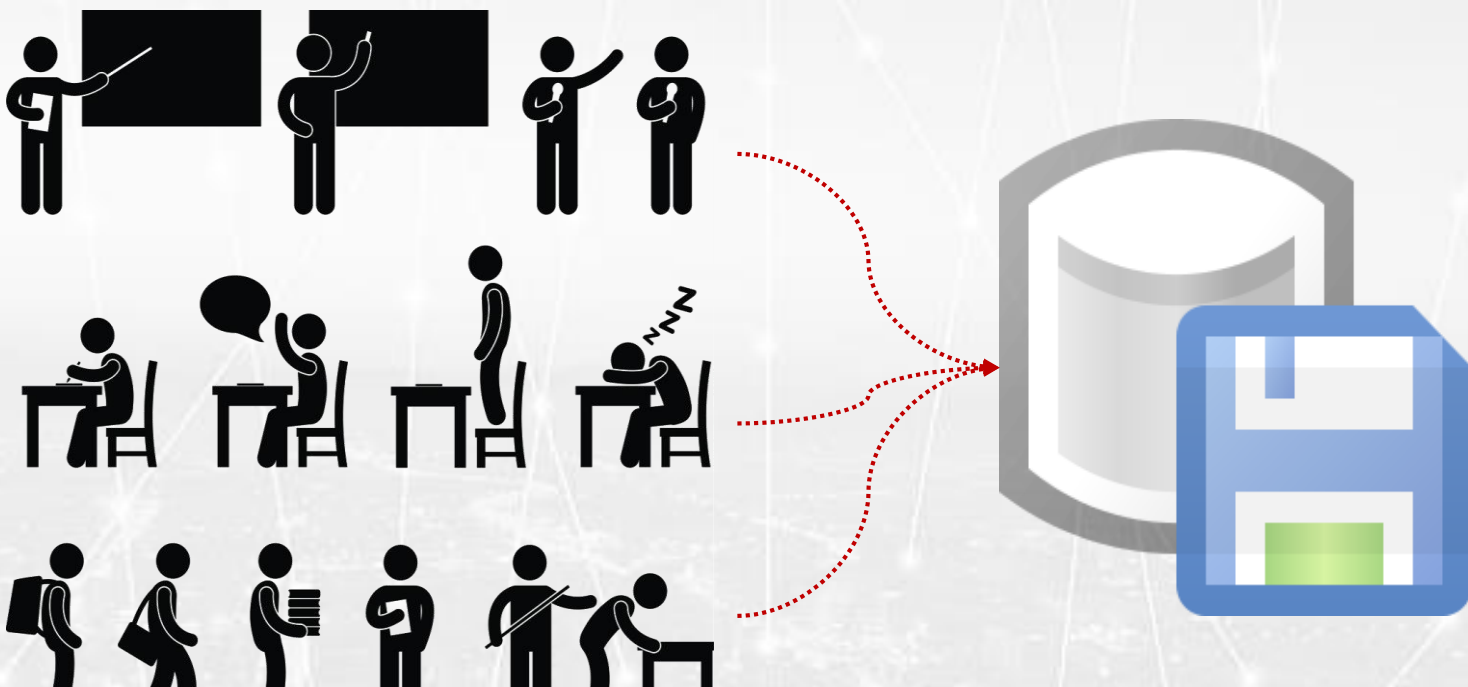
图书馆

对异常行为进行分析，探究缘由，针对性的分析和设计教学/学习方案

行为数据

➤ 大数据在高等职业教育中可以这么用

个性化教学



把学习者行为的数据放入数据库，根据数据调用不同分析工具和模型进行分析，从而**为学习者提供不同的内容和学习策略**，个性化指导学生学习。

➤ 大数据在高等职业教育中可以这么用

个性化教学

存到哪个数据库？

没有那么多精力来学习技术？

怎么收集数据？

有数据怎么分析处理？



➤ 大数据在高等职业教育中可以这么用

个性化教学



✓ 从小数据积累成大数据；

1节课的数据-1周的数据-1学期的数据.....

1名学生的数据-1个班级的数据-1个学校的数据

✓ 利用多种工具收集学生数据；

客观题答题正确率与问卷星、调研宝；

学习平台统计学生作业完成情况、学习时长等

➤ 大数据在高等职业教育中可以这么用

管理 精细化

□ 科研经费管理精细化

- 来源多元化：政府资助、学校资助、企业资助.....
- 粗放式管理容易出现管理断点、信息孤岛等问题

□ 考务管理精细化

- 在大数据理念下，在规范化的考务管理中快速有效找到需精细化的考务细节,使用信息技术，实现考务细节的制度化、规范化、科学化和精细化。

□ 学生工作管理精细化

- 用于思政教育，使学生行为、思想数据不再以孤立、碎片无序的形式存在；
- 用于学生就业指导；
- 用于心理健康监督及指导；
- 用于奖助学金等评定

➤ 大数据在高等职业教育中可以这么用

大数据与学生工作.思政教育



以前



网络学习时长

课堂教学活动互动次数

有关话题讨论的关注

.....

大数据

➤ 大数据在高等职业教育中可以这么用

大数据与学生工作·就业指导

利用大数据技术对学生进行职业生涯规划指导，帮助学生寻找自己的职业兴趣点与特长，以制定出较为符合学生身心健康发展和社会需求的职业生涯规划。



霍兰德职业兴趣测试

➤ 大数据在高等职业教育中可以这么用

大数据与学生工作.就业指导



在大数据分析的基础上，根据大学生职业规划的需求，在专业设置与课程选择上更具有灵活性，根据社会需求及时调整教学计划，进而使学生能够掌握适应未来工作与生活的知识技能。

➤ 大数据在高等职业教育中可以这么用

大数据与学生工作.心理健康监督及指导

帮助学生工作者深入了解所管理的学生，并能筛选出特殊群体和心理问题易感群体，使其及时给予关怀与指导，从而针对学生的心理问题做到早预防、早发现、早干预、早治疗。

学习行为

消费行为

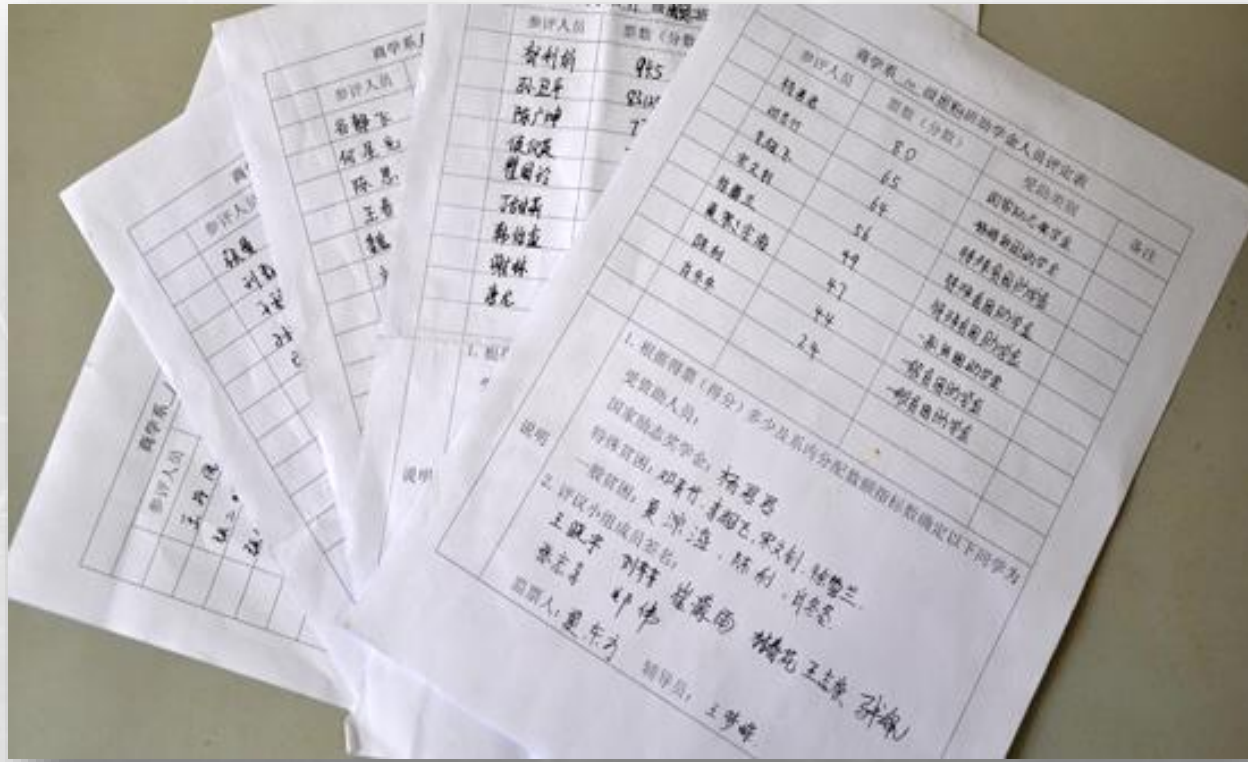
心理倾向

原生家庭环境

事实数据&行为数据

➤ 大数据在高等职业教育中可以这么用

大数据与学生工作.奖助学金评定



以校园卡所记录的数据为例，通过对参评者在校园内消费时间、地点及金额的统计，以了解他们的生活习惯，推测其在学校的消费水平。学生工作者将这些信息与他们的家庭经济情况、生源地、家庭成员构成、性别、年龄等信息进行关联分析，进一步了解到他们在校内消费水平与上述因素是否具有某种关联性，通过大数据分析技术可以建立全体学生的常模参数。



感谢聆听