

edX在线课程分析

摘要

1

背景

- edX是麻省理工和哈佛大学联合发布的一个网络在线教育计划。
- 分析的数据是该平台2012年9月-2016年8月290个在线课程的数据。希望通过分析，了解平台和用户的整体情况，探索优化产品、提高教学效果的方法。

2

平台生态

- 2013-2015年平台课程数量以每年近一倍的速度上升。
- 课程共分为4大类目，其中人文、历史、设计、宗教和教育类课程数量最多。两所学校各有所侧重。
- 2013-2015年累计报名人数呈持续上升趋势，增长率有所下降。2015年累计报名人数为156万+。
- 2016年1-7月累计报名人数同比下降明显。

3

用户生态

- 平台近七成学员为男性，各课程用户年龄中位数集中在22-38岁之间，67%的用户为本科及以上学历。
- 近些年热门的计算机类课程最受欢迎。
- 课程完成度和认证率均较低。
- 课程转发到论坛的比例约为8.2%。

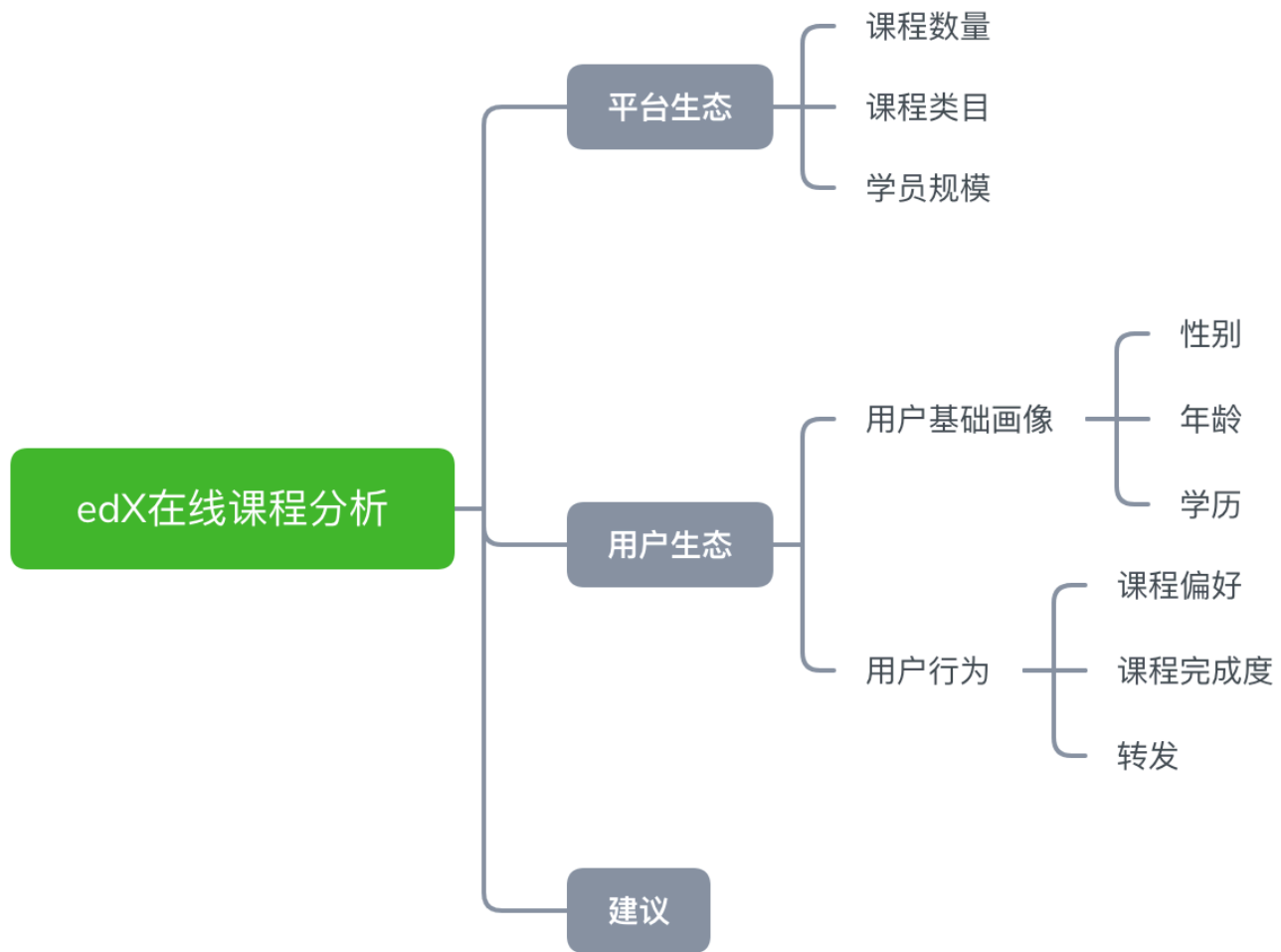
4

建议

- 增设计算机类课程，完善课程分类分级，不同机构在引入不同主题课程时优先级不同。
- 男性优先推荐理工类课程，女性优先推荐人文类课程。计算机和人文历史优先推荐Harvard的课程，卫生和社会学、工程和机械类课程优先推荐MIT的课程。
- 设置上课提醒、阶段性成果检测、开设学习小组等，提高学习完成度。

背景

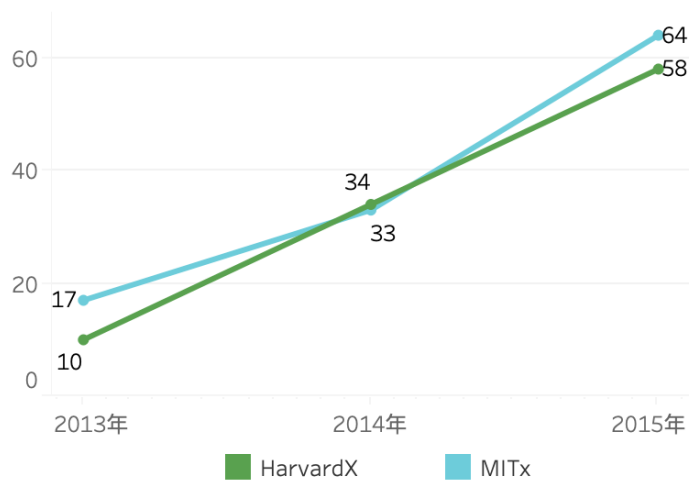
- 大规模在线开放课程，即MOOC (massive open online courses)。2012年，美国的顶尖大学陆续设立网络学习平台，在网上提供免费课程，出现了edX、Coursera、Udacity三大课程提供商。这三大平台的课程全部针对高等教育，且课程都是免费的。
- 本文分析的数据是edX上2012年9月-2016年8月MIT和Harvard在线课程的数据。包括290门课程、23个字段信息。
- 希望通过分析，了解平台和用户的整体情况，探索优化产品、提高教学效果的方法。



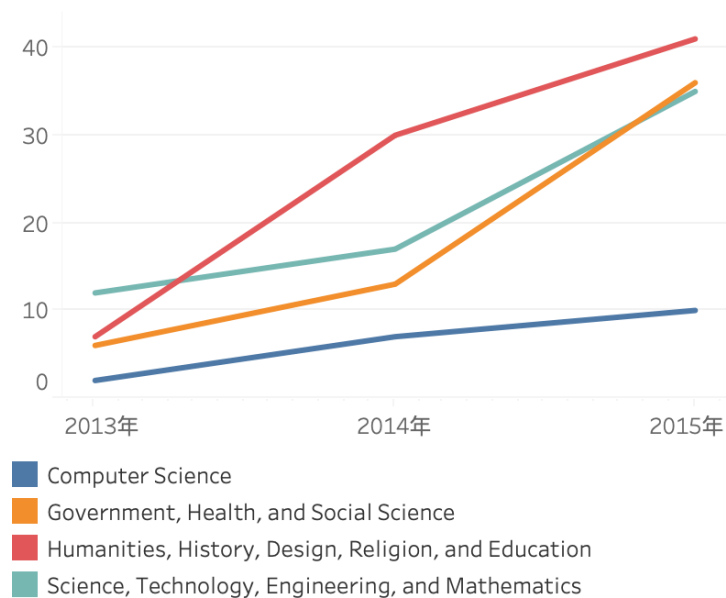
平台生态——课程数量

2013-2015年平台总课程数量、两校课程数量均以每年近一倍的速度上升。各主题课程数量也均为上升趋势。

2013-2015年两校课程数量走势



2013—2015年不同主题课程数量走势

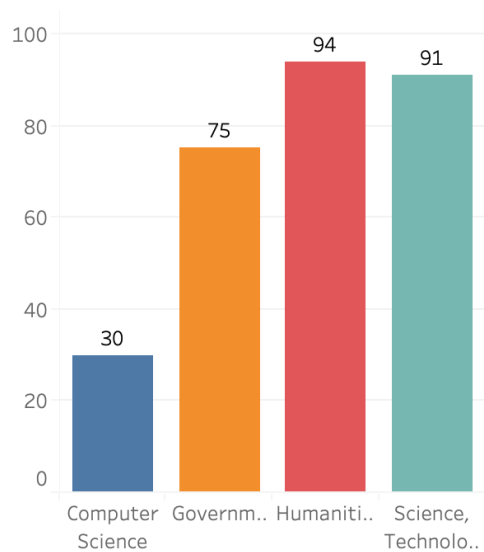


平台生态——课程类目

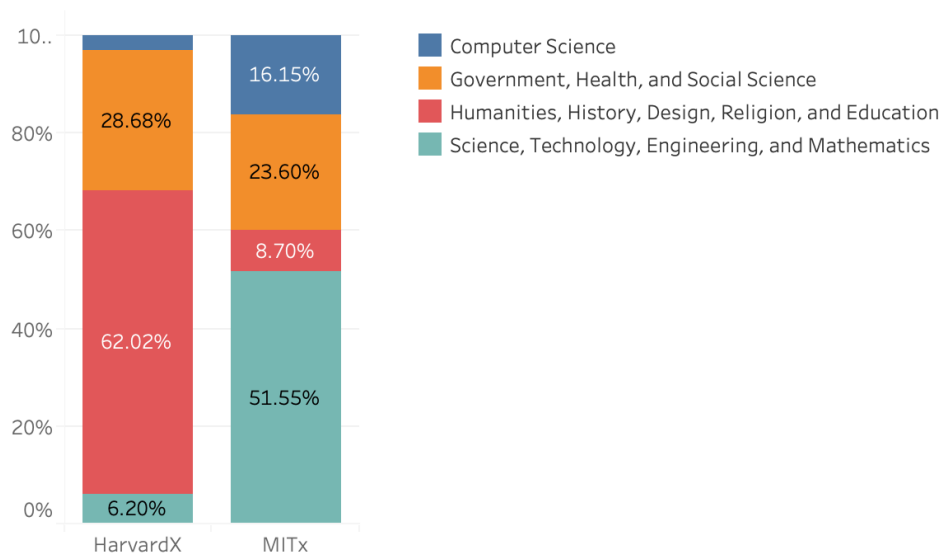
人文、历史、设计、宗教和教育类课程数量最多。两所学校的课程种类分布差异明显。

Harvard的人文、历史、设计、宗教和教育类课程占比最大，偏人文；MIT的科学、技术、工程和机械类课程占比最大，偏理工。

各类主题课程数量



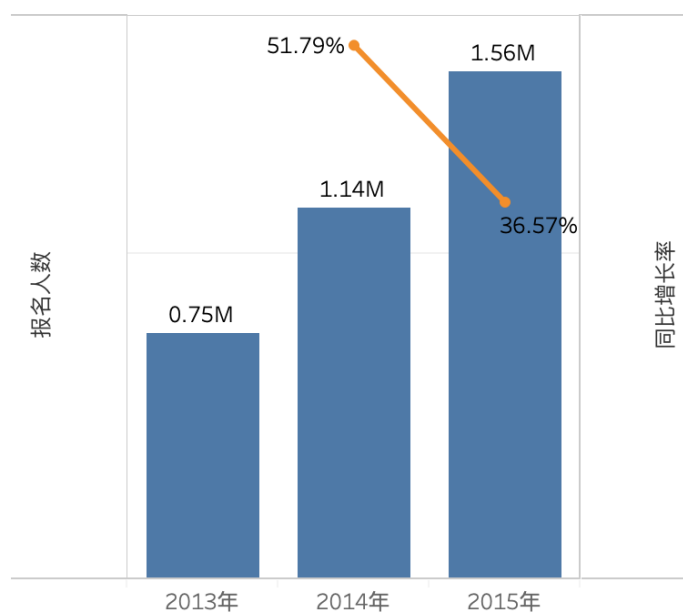
两校课程主题分布



平台生态——学员规模 (1)

2013-2015年累计报名人数呈持续上升趋势，增长率有所下降。

2013-2015年各年度累计报名人数及增长率

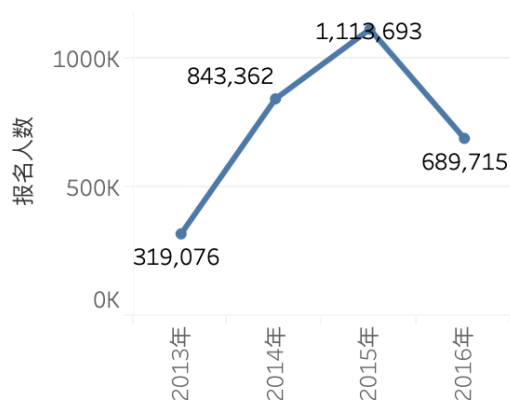


平台生态——学员规模 (2)

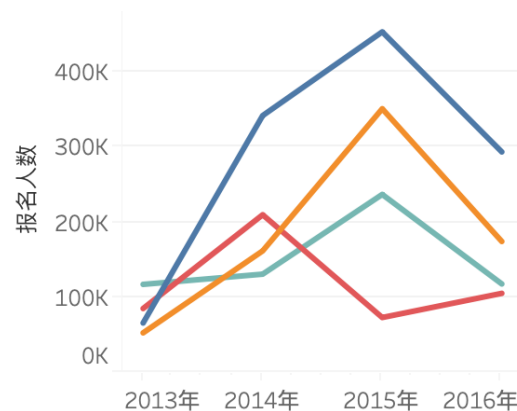
2016年1-7月累计报名人数有明显下降。

进一步拆解，各类课程报名人数基本上都有所下降，而课程数量基本是上升趋势，因此报名人数的下降需要结合当时市场环境、竞对平台情况、运营活动、下降群体等多方面进一步分析。

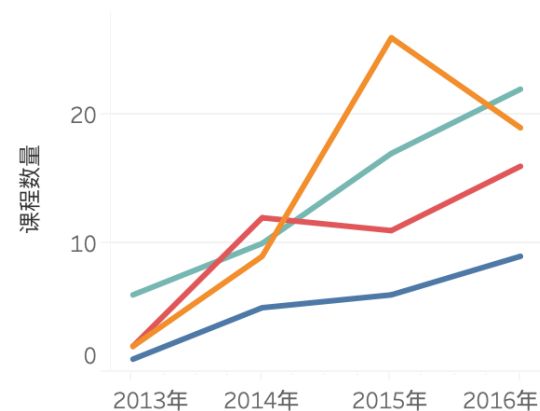
2013-2016各年度1-7月累计报名人数走势



2013-2016各年度1-7月不同主题课程累计报名人数走势



2013-2016各年度1-7月不同主题课程数量走势



- Computer Science
- Government, Health, and Social Science
- Humanities, History, Design, Religion, and Education
- Science, Technology, Engineering, and Mathematics

用户画像——性别

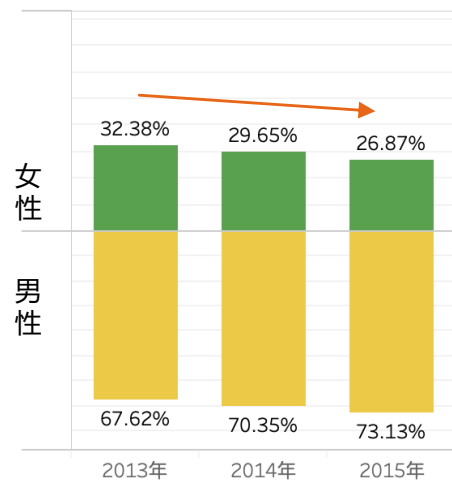
平台近七成学员为男性，各类目下学员性别分布有一定差异。

2015年平台整体男性学员占73%，女性学员占比连续两年下降。

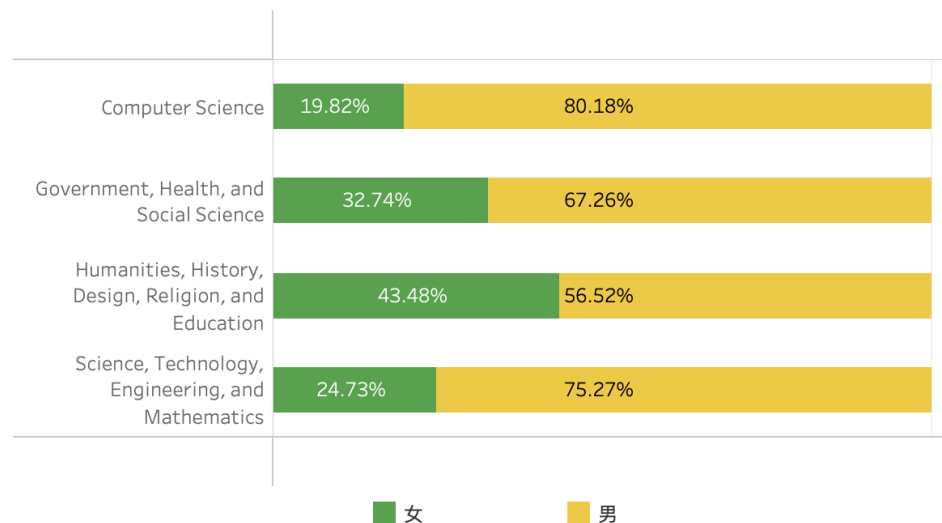
在计算机科学、科学·技术·工程和机械类课程主题下，男性学员占比远高于女性学员。

在人文、历史、设计、宗教和教育类课程下，男女学员占比相对均衡，男性学员占比略高于女性。

学员性别分布



各课程主题下学员性别分布



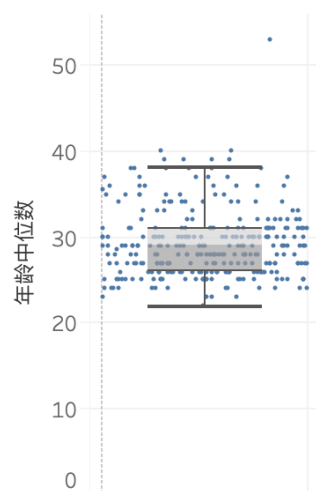
用户画像——年龄

各课程学员年龄中位数在22-38岁之间，不同课程主题下学员年龄中位数分布有一定差异。

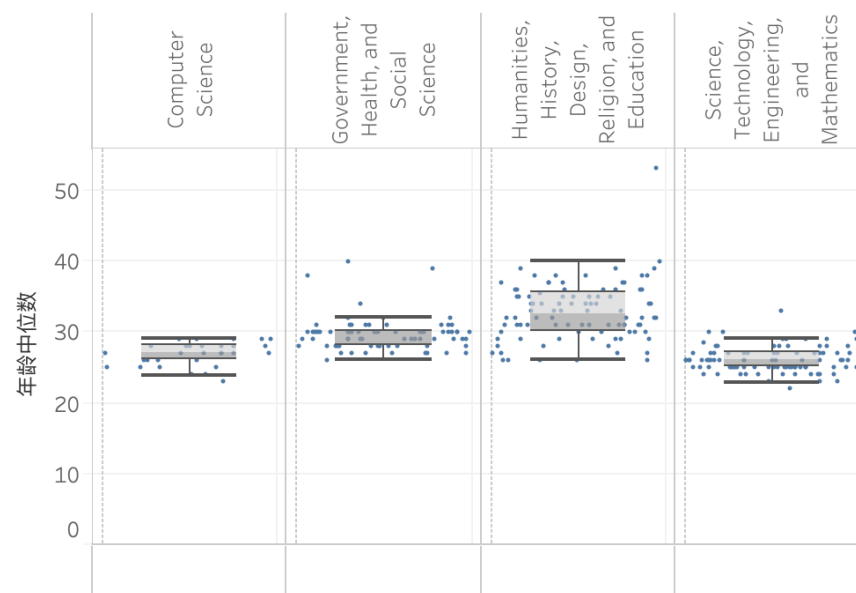
计算机科学和科学·技术·工程和机械这种偏理工类课程主题的学员年龄中位数相对较低，为23-29岁。

政府·卫生和社会科学、人文·历史·设计·宗教和教育这类偏人文类课程主题的学员年龄中位数相对较高，其中人文·历史·设计·宗教和教育的学员年龄中位数最高，为26-40岁。

各课程学员年龄
中位数分布



各课程主题下学员年龄中位数分布



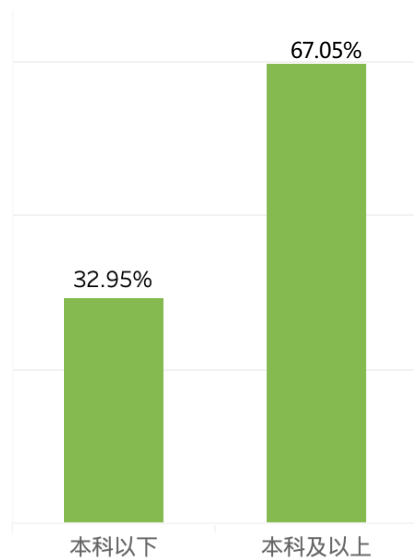
用户画像——学历

平台整体近七成学员为本科及以上学历，各类目下学员学历分布有一定差异。

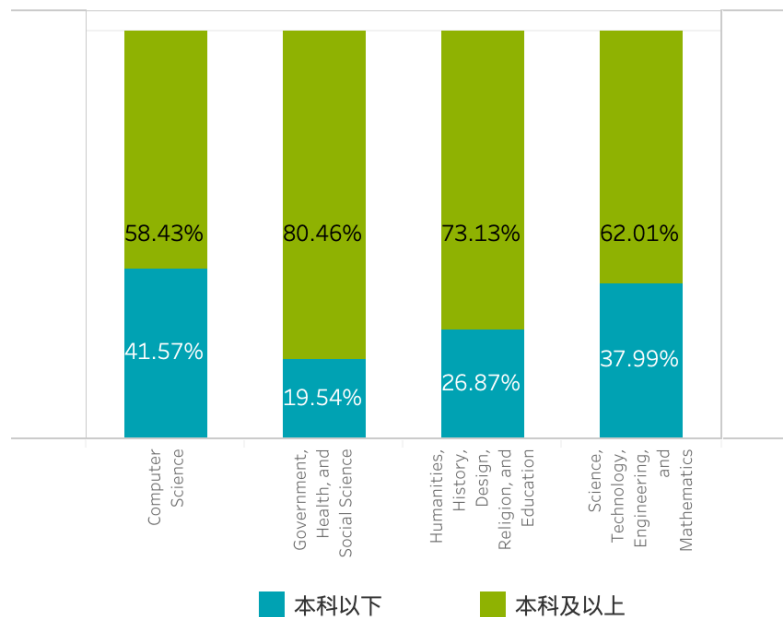
本科及以上学历学员占67%，这和平台上课程针对高等教育相符。

在计算机科学主题下，本科以下学历学员占比相对最大。

学员学历分布



各课程主题下学员学历分布

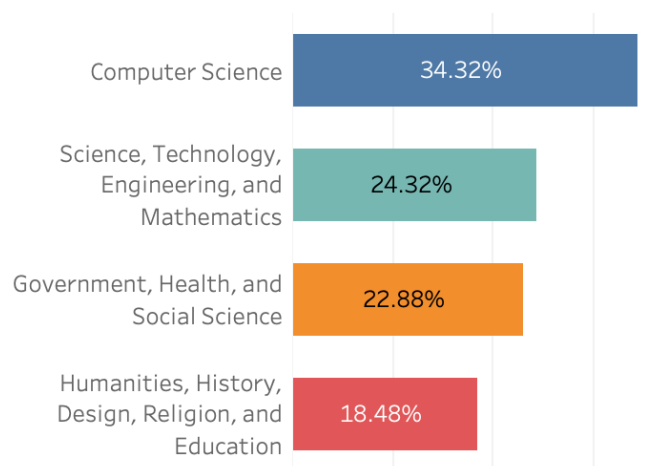


用户行为——课程偏好 (1)

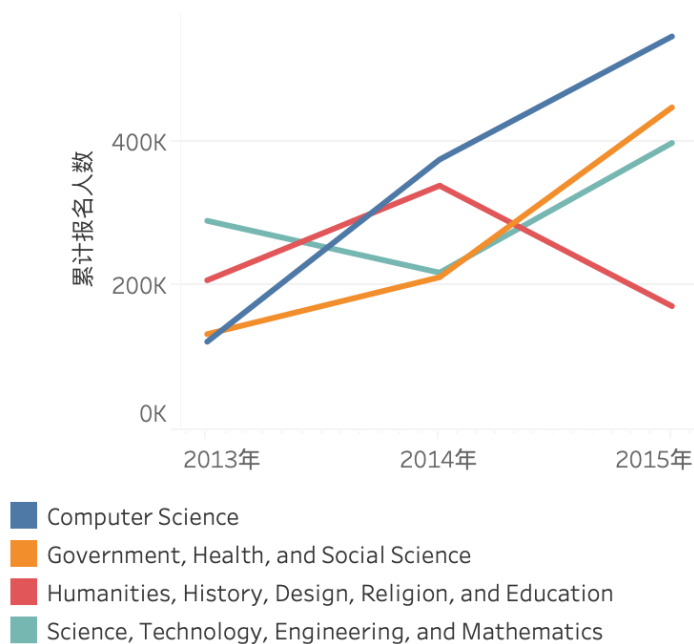
计算机科学类课程报名人数最多，且连续两年报名人数持续快速上升，与计算机科学的快速发展相匹配。

计算机课程需求最大，结合平台各主题课程数量，计算机科学类课程最少，应考虑开发计算机类课程。

各课程主题累计报名人数占比



各年度不同主题课程报名人数走势

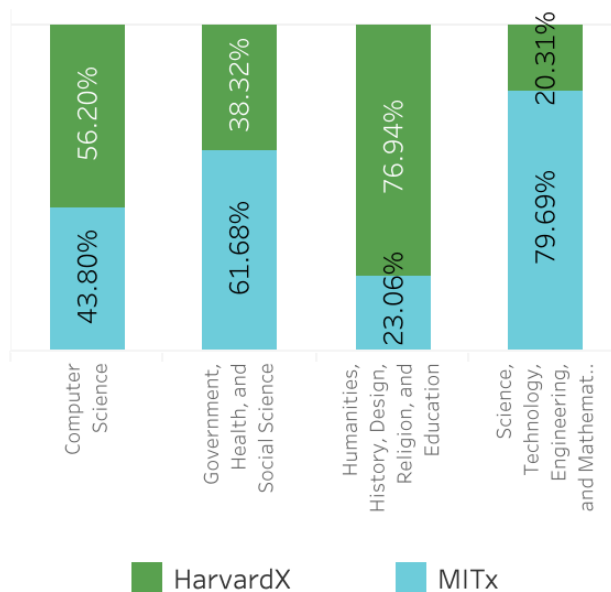


用户行为——课程偏好 (2)

各主题课程学员在两个学校的分布有差异。

在计算机科学和人文、历史、设计、宗教和教育类目下，Harvard的课程学员更多；政府、卫生和社会学，科学、技术、工程和机械类课程则是MIT的课程学员更多。

各主题不同机构学员分布

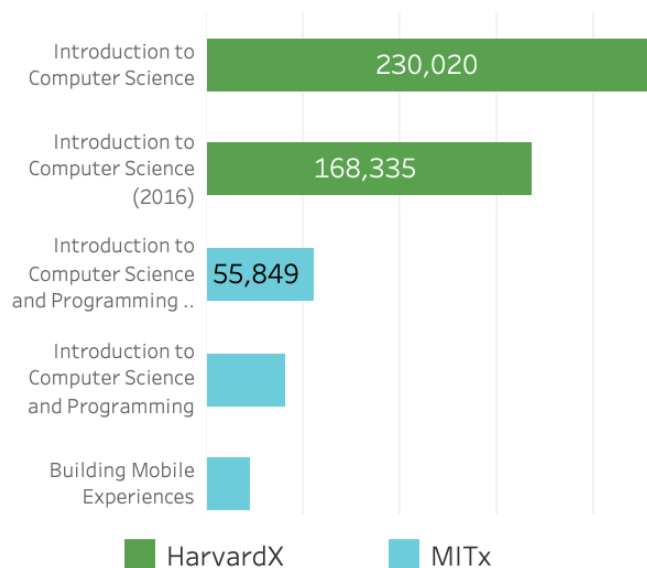


用户行为——课程偏好 (3)

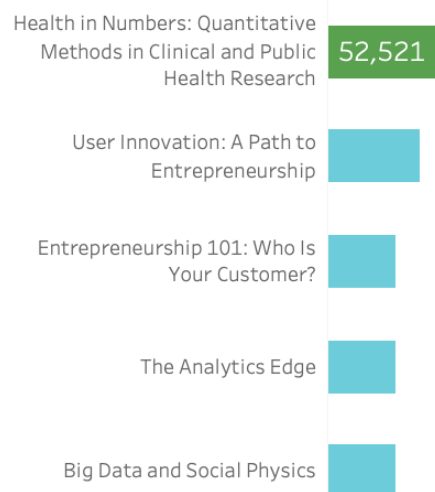
在计算机科学和人文、历史、设计、宗教和教育类目下，Harvard的课程更受欢迎；政府、卫生和社会学，科学、技术、工程和机械类课程则是MIT的课程更受欢迎。

各主题最受欢迎课程top5及来源机构

计算机科学



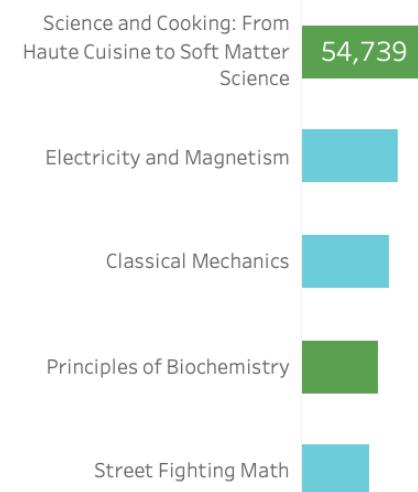
政府、卫生和社会科学



人文历史、设计、宗教和教育



科技、工程和机械

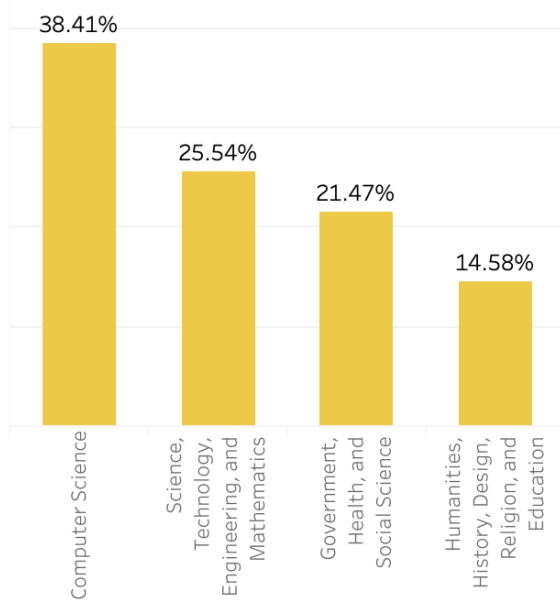


用户行为——课程偏好 (4)

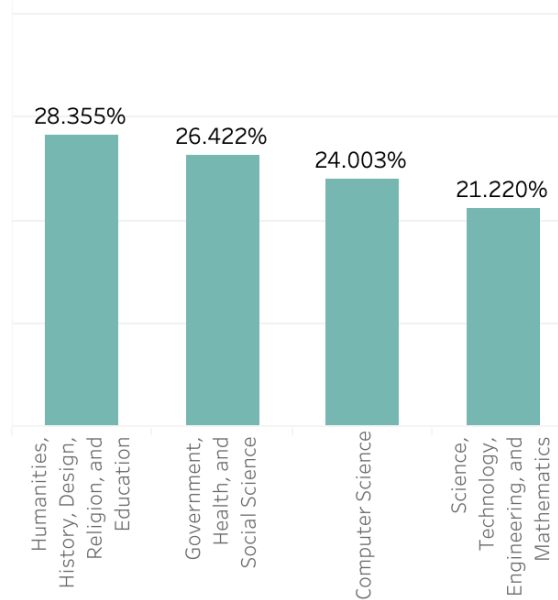
不同性别倾向的课程类别差异明显。

男性用户更偏爱计算机科学，科技、工程和机械类课程，偏理工；女性用户更倾向于人文、历史、设计、宗教和教育类课程。

男性在各主题分布



女性在各主题分布

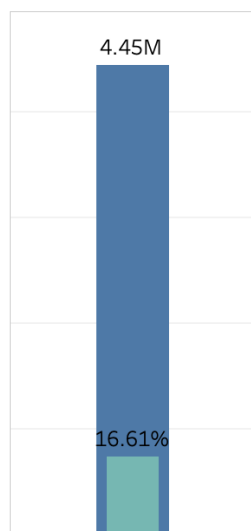


用户行为——课程完成度 (1)

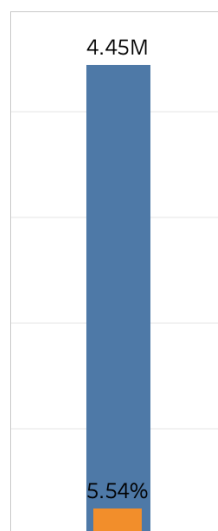
平台报名人数很多，但完成过半课程内容的人数占比和获得认证的人数占比都很低。
计算机科学类课程完成过半课程内容和最终获得认证人数占比是几类课程里最低的。

可以认为坚持完成大部分课程内容及最终获得认证的学员，学习目标是比较明确的。绝大部分学员则无法坚持。
有很大一部分学员受到近些年计算机行业快速发展的影响而对计算机课程感兴趣，因此这类课程学员人数多，但完成度低。

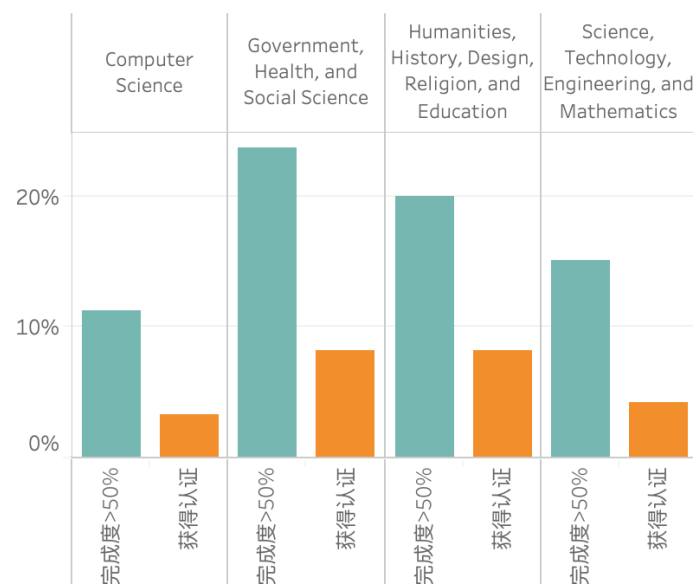
完成>50%课程
人数占比



获得认证
人数占比



不同主题课程完成度



注：总分达到最低分数要求时会获得荣誉证书。

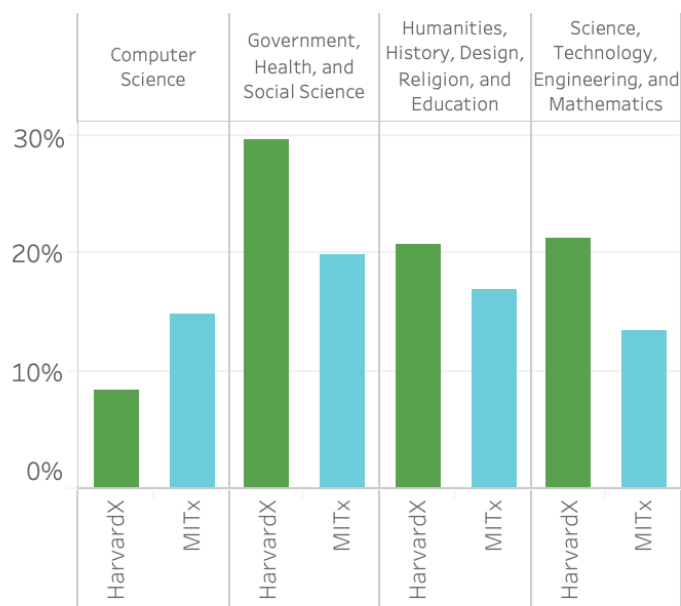
用户行为——课程完成度 (2)

不同机构同类课程完成>50%内容人数占比和认证人数占比趋势一致。

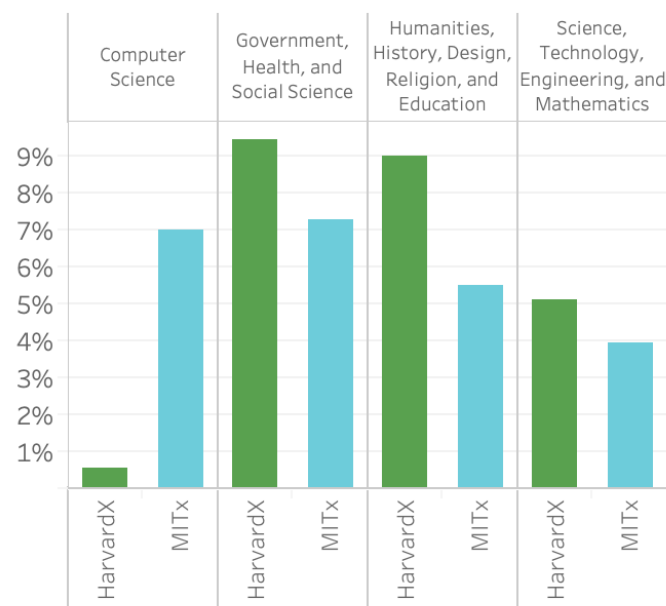
除人文、历史、设计、宗教和教育主题外，其他主题课程报名人数相对多的机构课程完成度和认证率都相对更低。

一可能是课程内容较难；二可能是因为用户选课时会受到已报名人数影响，或者课程列表内报名人数多的课程排名更靠前，更多的学员因此报名学习，而没有认真考虑课程是否满足自己的需求，导致报名人数占优势的机构课程完成度反而低。

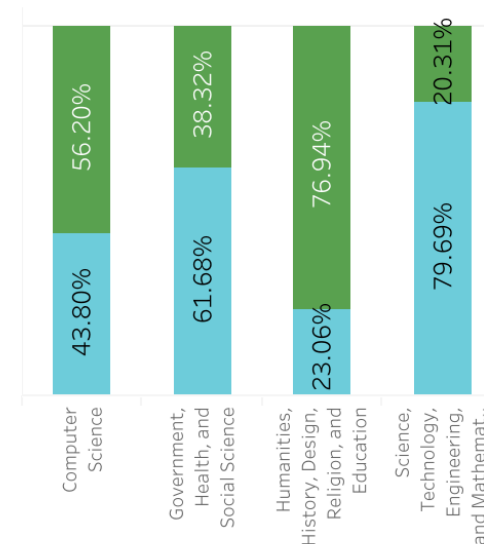
不同机构同类课程完成>50%占比



不同机构同类课程认证率



同一课程主题不同机构学员分布



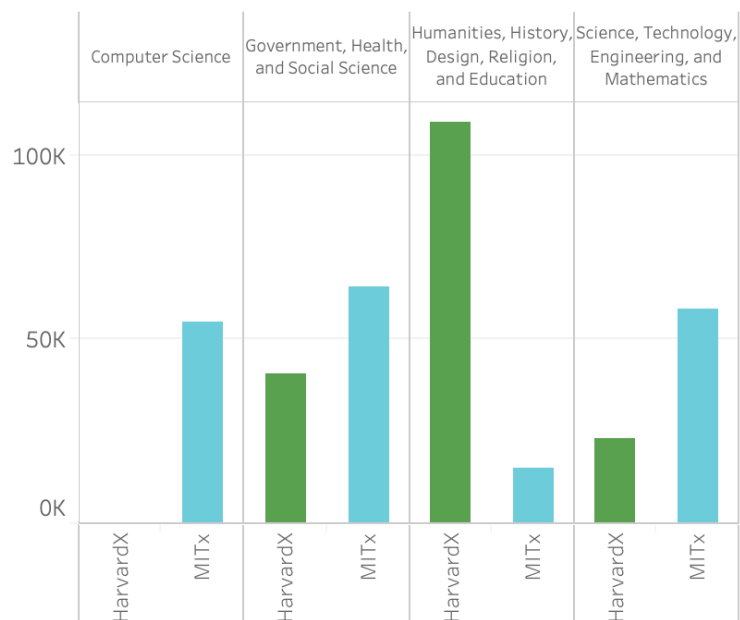
用户行为——转发

Harvard的人文、历史、设计、宗教和教育主题课程转发到论坛的数量和占比都最高。

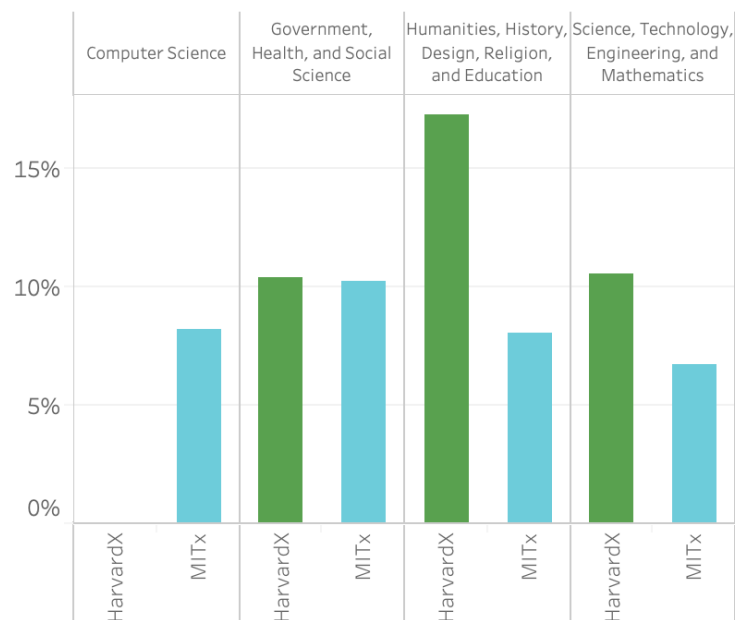
发布到论坛的数量可反映课程的受关注程度。Harvard的人文、历史、设计、宗教和教育类课程是该校课程数量最多的种类，且转发次数和占比都远高于其他课程，表明其课程质量确实很高。

Harvard的计算机科学类课程报名人数很多，但转发为0，查看数据发现该校计算机类只有计算机科学一门课程，开课4次均没有转发，需要查看是否数据源有问题，或者课程页面没有转发按钮。

不同机构各主题课程发布到论坛数量



不同机构各主题课程发布到论坛数量占比



建议

1

课程开发

1. 引进课程时，计算机科学和人文、历史、设计、宗教和教育类目，优先考虑Harvard的课程；政府、卫生和社会学，科学、技术、工程和机械类课程优先考虑MIT的课程。
2. 增加计算机科学类课程数量。
3. 细化课程分类、完善课程难易程度分级、完善课程介绍，方便用户更快速准确地选到符合自己需求的课程。
4. 长期来看，需要调整课程内容提高课程质量。

2

课程推荐

1. 用户登录未搜索课程时，男性用户优先推荐Harvard计算机科学和MIT的科技、工程和机械类课程；女性用户优先推荐Harvard的人文、历史、设计、宗教、教育和MIT的政府、卫生和社会科学类课程。
2. 在用户进入特定课程主题或搜索过特定主题的课程时，计算机科学和人文、历史、设计、宗教和教育类目下，优先推荐Harvard的课程；政府、卫生和社会学，科学、技术、工程和机械类课程类目下优先推荐MIT的课程。
3. 收集更详细的用户信息，如年龄、学历、专业、行业、地域、收入等，从不同维度探索不同特征用户的课程偏好，完善个性化推荐。

3

学习设置

1. 设置上课提醒、阶段性成果检测、开设学习小组等，强化学员的参与感和成就感，调动学习积极性，提高学习完成度。
2. 收集课程反馈，用于改进课程。