## 前言

在行是果壳网于2015年3月推出的一款内容中介产品。在在行上，用户可以根据自己关心的问题，挑选在相关问题上有能力、有经验的“行家”，一对一见面咨询，进而帮助用户解决问题。

在行的模式在推出之时取得了不错的反响。在行及在行团队后来孕育出的分答，在一定程度上引领了国内知识付费的浪潮。而时至今日，在行推出两年多之后，热度却渐渐消退。

在行的发展轨迹，以及这几年知识付费领域的起起伏伏，其中的道理值得探究。本篇报告，我们将从爬虫爬取的在行数据出发，以数据分析为基础，以文字分析为脉络，描绘出在行这个产品的轮廓及特征。之后，再以在行这个典型的产品为样本，尝试讨论知识付费这个领域的一些问题。最后，我们还可以经由之前的分析和讨论，对整个信息传递这样庞大的领域有一些判断和理解。

本篇报告将从产品、市场、供给、需求和运营五个方面来对在行进行分析。

以下是报告的目录：

## 1 在行产品分析

本章着重介绍在行作为一个互联网产品在运营上的基本情况，并尝试结合数据，对在行做出一些总结和描述。

### 1.1 在行简介

在行于致力于为用户提供精准的专业内容服务。官方的自我定位是“知识技能共享平台”。

在行上有两类人，一类是“行家”，是内容的提供者；一类是普通用户，对行家进行约见，是内容的需求方。

行家在在行展示技能或经历，发布话题；而用户则选择中意的话题，约见行家，通过不同的方式向行家学习、请教，或与行家交流、沟通；进而实现问题的解决，信息、知识、技能、经验的传递或交流。

而如果完成了一次上述的流程，我们则称，完成了一次约见。

约见完成之后，用户会选择对约见的话题做出评论和评分，其余人可以对评论点赞，而行家也会对一些评论做出回复。

有时，用户会关注他们想约见的行家，将之加入心愿单。如此，行家信息中的“想见”数便会增加。用户可以借此提醒自己，以便日后合适之时进行约见。

在行能够让许多普通人也能接触到行业专家或知名人士，并进行近距离沟通交流。在行的这项功能降低了知识传播的门槛，提升了知识传播的速度和广度。

在行为行家的入驻提供认证和筛选，行家于在行上展示的个人经历或话题内容，都经过考量和设计，以求能更好的展示出行家的专长，同时吸引合适的用户前来约见。

在这种情况下，在行实际上提供了对“内容”这一商品的信用背书，将散乱、模糊的各种沟通形式与环节统一化、透明化。省去了寻找合适的专家，打探联系方式等诸多麻烦。

内容服务的核心环节，在于双方的交流沟通。无论是否通过在行，这一环节的形式和内容都不会有什么区别。但在行的出现，一是能够大大提高见面前后环节的效率，二是能够促使很多原本不能发生的交流沟通得以实现。从社会角度讲，在行能够提高内容的传播效率和传播效果。考虑到在行对内容的关注偏向于专业性，实用性，在行有着提高社会生产效率和生产力的潜在可能性。

### 1.2 数据概览

报告所采用的数据大部分采集于2018年1月初，并于2月下旬进行了更新和补充，最终抓取在行行家数据8228条，话题数据14976条，评论数据115564条，用户数据53012条。

借此，我们可以大致地了解在行。

在行于2015年3月13日推出，而我们的数据最早记录到了2015年1月。截至2018年1月5日，在行共有行家8228名，话题14976个；发生约见158477次，产生评论116614条，评论率为73.58%，这其中又有25447条评论被行家回复，回复率为21.82%；累计付费用户约72697名（修正），并产生了865136次对行家的关注行为。

在此期间，在行上的累计成交额约为97540413元（修正），平均约见价格为615元。

从这些数字，可以看出，在行的规模偏小，在行上约见的数量和频率偏低。但是在行的客单价和用户价值较高，行家和用户之间的交流积极，用户反馈度高。

这便是在行的基本情况，接下来，本报告将会从不同方面、不同角度来对在行及在行的数据进行解读分析。

### 1.3 关键指标

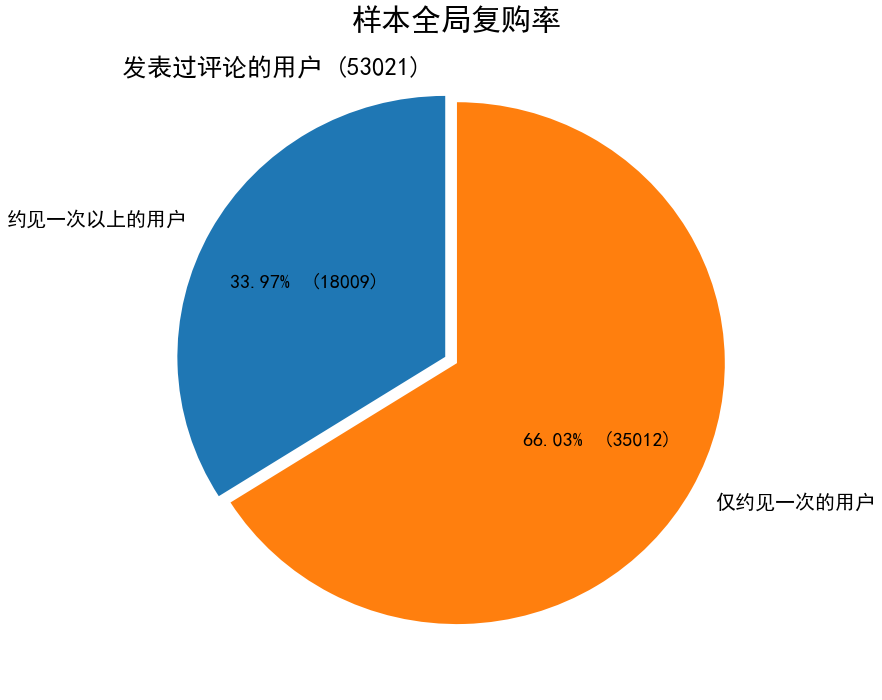
在运营上有一些重要的考核指标，如点击率、增长率、留存率、复购率、转化率等等。本节将对其中几个指标进行计算和分析。

#### 1.3.1 样本全局复购率

复购率考核消费者对产品的重复购买情况，假设某产品共有10个用户购买，其中5个用户购买了一次以上，那么复购率为50%。

消费者的重复购买率越高，则对产品的忠诚度越高。复购率侧面反映了用户对产品的满意程度、需求程度或依赖程度

关于复购率，针对不同的标准和周期可以计算出许多不同的比率。由于在行上的约见属于低频率，长周期的消费行为，我们先来看一看，至今为止共有多少用户进行了一次以上的约见。



如图所示，对约见做出了评论的用户共有53021位，其中18009位进行了重复购买，那么在行的全局复购率为33.97%，属于良好水平。

考虑到有26.4%的约见并没有被用户评论，那么这部分用户应该属于对约见满意度和积极度较低的类别，因此在行真实的全局复购率应该要更低一些。

#### 1.3.2 关注-约见转化比

在在行上，如果想要约见某位行家，却因为一些原因不能立即行动，那么可以将这位行家加入心愿单，留待以后约见。点击了“加入心愿单”，就发生了一次关注行为，行家信息上的“想见”人数也会相应增加。



将行家加入心愿单这个行为，需要登录在行，因此这项数据的可靠性较高，具有一定的参考价值。

如果将在行约见的流程整理为：进入在行-挑选话题-联系行家/加入心愿单-商谈约见事宜-完成约见-评论-回复。可以看出关注行家的人数是约见之前一个重要的流量数据。因此关注-约见转化比可以作为衡量在行付费转化率的一个指标。

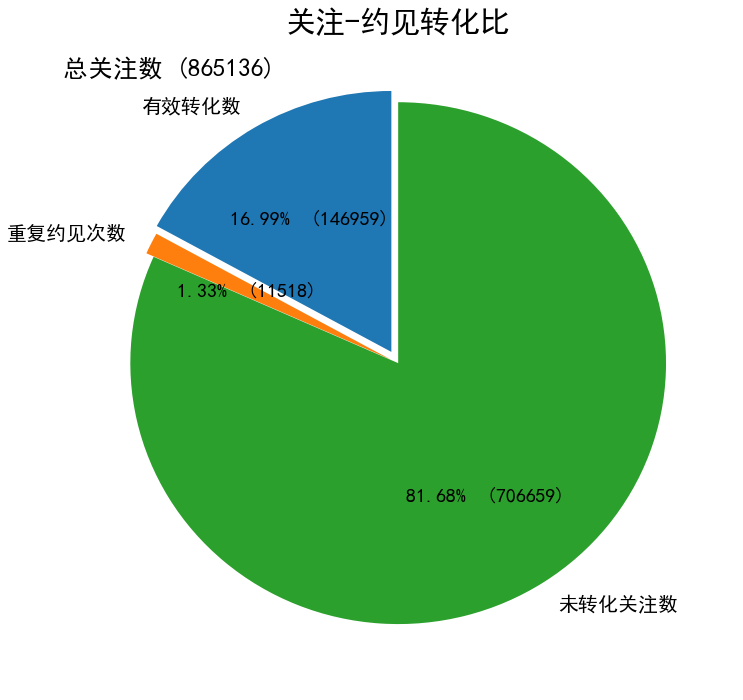
转化率广义上指某个用户群体中，通过商家的影响，最终完成了商家所期望行为的用户所占的比例。针对在行，就是接触到在行的用户中，通过了解、流量在行，最终在在行上进行了约见的用户所占的比率。

转化率与复购率一样，根据不同的标准有不同的计算方式。

首先我们假设所有的约见者都事先关注了行家。

因为同一用户对两个行家的关注是算作两次的，那么，同一用户对不同行家的约见应该算作多次转化。因此在对约见进行去重时，将只除去同一用户对同一行家的重复约见。

在抓取的115564条评论中，有107165条是不同用户与不同行家的约见。将这个数字放大至总约见次数158477次中，为146959次。



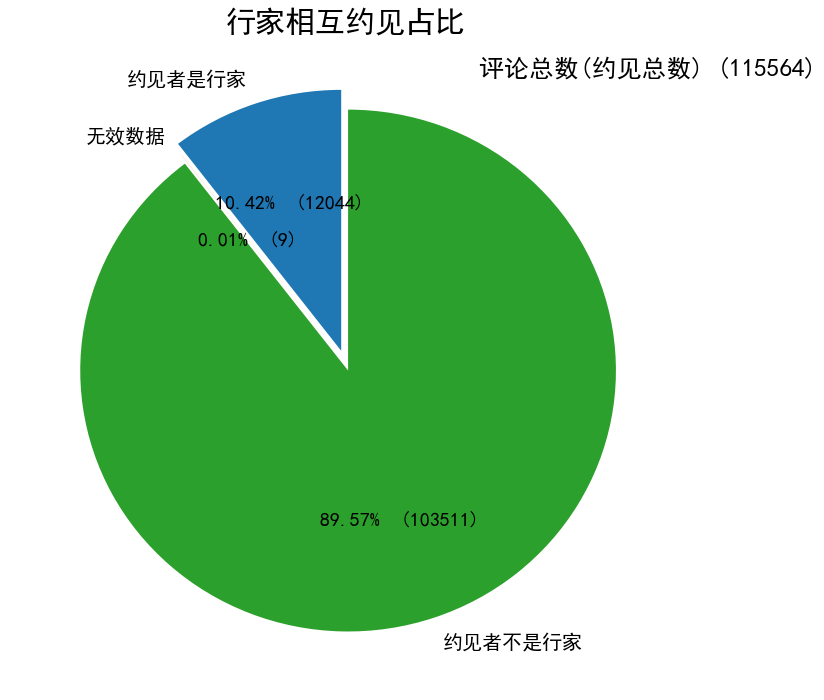
如图，转化率为16.99%。因为关注是距离约见非常近的环节，相当于传统电商中收藏商品或把商品加入购物车，所以这个转化率不算高。但是考虑到在行约见的周期、约见困难，出现较低的转化率也是正常的。

总之，因为在行产品的创新性和独特性，我们很难断言在行的转化率是好是坏。如果对这个指标进行长期跟踪，监控转化率的变化，并执行相应的对策，才能体现出它的作用。

#### 1.3.3 行家互相约见占比

这是在行独有的一个指标，在行作为一个新颖的产品，在发展初期，许多行家和用户都是在好奇心和新鲜感的驱使下来到在行。有许多行家，体验过被用户约见之后，也想尝试约见别的行家。这样的情形究竟普遍吗？

以下数据来自抓取到的115564条评论，在此作为总约见数对待。



有10.42%的约见是在行家之间进行的，这个比率大致等于行家与用户总数的比率：8228/ 72697=11.32%。但是这样的对比是十分不严谨的。如果仔细研究这11万次约见的用户身份，会得到这样的数据：

|  |  |  |  |
| --- | --- | --- | --- |
|  | **评论者** | **约见者** | **人均约见** |
| **不是行家** | 103511 | 50309 | 2.0575 |
| **是行家** | 12044 | 2703 | 4.45579 |
| **无效** | 9 | NaN | NaN |
| **和** | 115564 | 53012 | NaN |

总数为115564次的约见，是由53012位用户完成的。其中2703位约见者同时也是在行的行家，虽然他们只占用户总数的5.1%，但是却贡献了11.64%的约见。行家约见者的人均约见次数达到了非行家约见者的两倍以上。

值得注意的是，这些行家占行家总数的32.85%。在上文中提到，用户总体的关注-约见转化比为16.20%，而付费用户只是全天注册用户的一部分，那么付费用户占全体注册用户的比例则要更小。如此对比之下，行家群体成为付费用户的比例显然高于非行家用户。

这些都说明，行家这一群体对在行确实有着很高的积极性。而他们如此活跃的原因，根据在行相关的资料文章，是因为他们有着更强烈的好奇心和探索欲。

这项指标的意义，在于帮助了解在行用户群体的分化，以及行家群体的特征和活跃度。由于在行上的约见是双方互动的行为，因此任意一方的表现对在行来说都很重要。

#### 1.3.4 总结

在行的全局复购率表现良好，说明在行解决了一些痛点，有一定吸引力；但在行总体的付费转化较低，应该是因为约见的流程长，限制因素多；在行上行家的积极性比较高，因为行家站全体用户的比例很少，因此保持行家的高活跃度，才能保持供需的平衡和约见的顺利进行。

### 1.4 流量分析

流量对于互联网产品的生存和发展至关重要。说到流量，首先想到的是网站的UV（独立用户）和PV（访问量），但是这两项数据十分难以获取，要么数据模糊，要么范围不全面。

本节的流量分析，主要是从抓取的评论数据的日期信息为基础，经过加工和整理，来对在行的几个流量指标进行估计和分析。

本节将从约见成交数量，评论数量以及新增付费用户数量三个方面来对在行的流量进行一些初步的分析。由于时间信息的来源仅有评论这一个，因此本节的流量分析局限性很大，仅供参考。

首先在百度指数搜索在行，得到这样一条曲线：

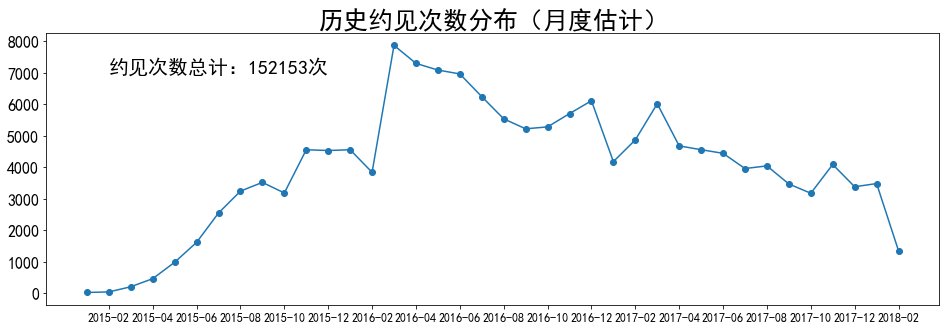


可以看出2016年6月之前，在行的热度一直在增加，之后则一直在缓慢下降。

百度指数的数据只是一个估计，且没有具体的含义。下面我们就从抓取的数据中来看一看在行的历史流量信息。

#### 1.4.1 历史约见次数的估计

根据话题信息和抓取的评论数据，可以大致估计出在行至今为止每月完成的约见次数。



这些数据是将每条评论按照日期排序分类，再按照约见乘数计数加总而来。

约见乘数是话题约见数与评论数的比值，如果某话题有10次约见，而有5条评论，那么约见乘数为2，意思是一条评论代表着发生了两次约见。假如在2016年1月这个话题有一条评论，便认为这个月此话题发生了两次约见，2016年1月的约见次数加两次。

这里假设了某个具体话题的评论率在任何时期都相等，即等于总体评论率。同时还假设了约见与评论之间的平均时间间隔保持不变。可以看出，约见数量统计的时间段越长，这样的估计方式准确度越高，为了兼顾准确度和参考价值，选择以月为单位进行统计。

途中估计出的总约见次数为152153次，与上文中统计出的总约见次数158477相差3.9%，属于可接受范围。而估计数比统计值少的原因是，有一部分评论相对应的话题已被删除，无法计算约见乘数，仅算作1次约见。

可以看出，在行在推出之后的一年内，每月约见次数呈快速上升的趋势，直到2016年3月达到顶点，约为7864次。之后便开始缓慢下降，到2017年12月约为3377次，不足顶峰的一半。其中在2016年3月、2016年12月、2017年3月，出现了几次尖峰，可以重点关注。

这条曲线的整体走势与百度指数大致相同，互相印证了准确度。

我们也可以借此反推在行的流量规模，以约见次数最多的2016年3月为例，如果1000个访客中，有1个进行了约见，那么在行的月UV为786万。当转化率为万分之一的时候，在行的月UV将达到7864万。在行的流量规模应属于中等偏上。

详细数据见下表：

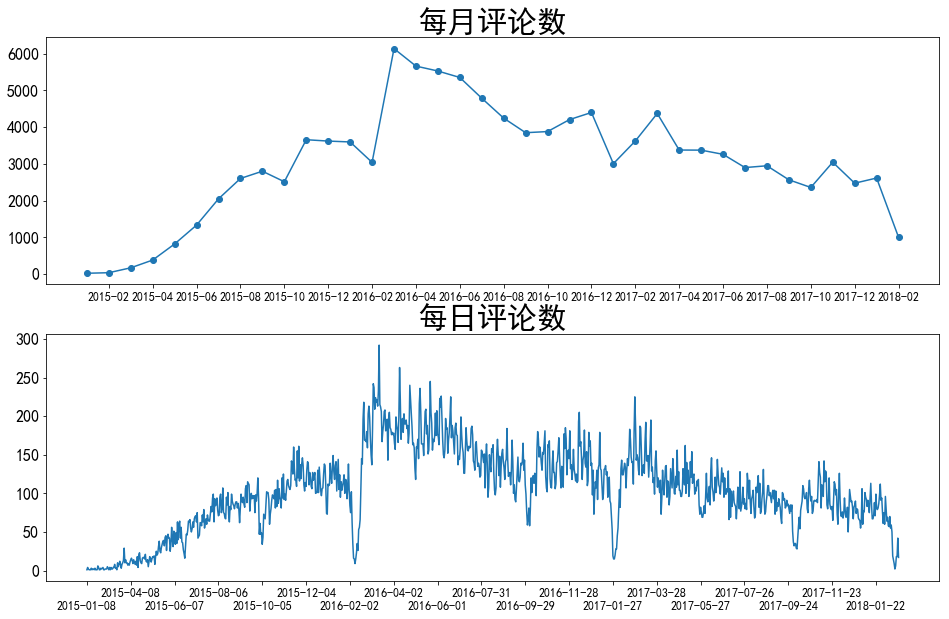
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **月份** | **约见次数** | **月份** | **约见次数** | **月份** | **约见次数** | **月份** | **约见次数** |
| **2015年1月** | 23 | **016-01** | 4551 | **2017年1月** | 4175 | **2018年1月** | 3479 |
| **2015年2月** | 40 | **2016年2月** | 3836 | **2017年2月** | 4864 | **2018年2月** | 1345 |
| **2015年3月** | 204 | **2016年3月** | 7864 | **2017年3月** | 6012 |  |  |
| **2015年4月** | 460 | **2016年4月** | 7292 | **2017年4月** | 4674 |  |  |
| **2015年5月** | 981 | **2016年5月** | 7081 | **2017年5月** | 4553 |  |  |
| **2015年6月** | 1613 | **2016年6月** | 6957 | **2017年6月** | 4438 |  |  |
| **2015年7月** | 2544 | **2016年7月** | 6237 | **2017年7月** | 3955 |  |  |
| **2015年8月** | 3239 | **2016年8月** | 5533 | **2017年8月** | 4039 |  |  |
| **2015年9月** | 3516 | **2016年9月** | 5216 | **2017年9月** | 3466 |  |  |
| **2015年10月** | 3180 | **2016年10月** | 5277 | **2017年10月** | 3171 |  |  |
| **2015年11月** | 4551 | **2016年11月** | 5698 | **2017年11月** | 4084 |  |  |
| **2015年12月** | 4527 | **2016年12月** | 6106 | **2017年12月** | 3377 |  |  |

在行是在2015年3月正式推出的，但是在1月和2月就已经发生了几十次约见，最早的记录在2015年1月12日。

我们也可以观察到，在行的约见次数在某几个月比相邻月份有大幅下降。在观察了详细数据后，发现这些下降是由于节假日的影响造成的。在行的约见次数，在春节、十一、五一等节假日均有显著下降，有时甚至跌至个位数。这说明在行的约见对行家和用户来说都是很严肃的事务，被当成工作而不是休闲娱乐对待。

#### 1.4.2 评论时间分布

根据抓取的评论数据，可以统计出评论时间的分布情况。

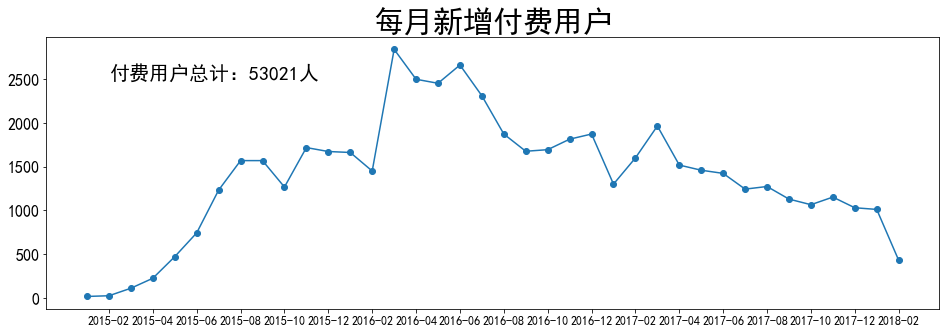


如图所示，在行每月评论数的分布曲线与每月约见数几乎完全相同，这是因为我们之前假设在行的评论率一直保持不变。这种假设的局限性很大，之后将尽量避免使用。

由于评论有详细的日期数据，我们可以统计出评论数量的日分布曲线。在这条曲线上，每年春节及十一期间出现的下降十分明显。同时也能看到，在其余非节假日的时段，也有许多周期性的尖峰和低谷。

#### 1.4.3 每月新增付费用户

最后再来看另一个时间分布曲线。



这是每月新出现的付费用户的分布曲线。他与评论及约见分布有极高的相似度，这说明在行的约见数与新用户的数量有极高的关联度。新用户多，当月约见数便相应提高，反之则下降。这也说明在行老用户并没有长久的留存下来，每月稳定的产生约见。否则约见曲线在后半段应当下降的很平缓。

#### 1.4.4 总结

在行在2016年3到达到流量顶峰，距离网站推出仅有一年，之后开始逐渐下降，增长期过于短暂。这个现象需要从多方面寻找原因，如曝光渠道、营销活动、产品迭代、市场特性等。在行的约见行为，在节假日会显著下降，一般来说节假日是产品运营推广的好时机，在行的这个特性可能会对其发展产生影响。在行的约见数量与新用户数量高度相关，这说明老用户的长期留存率和稳定程度不好。

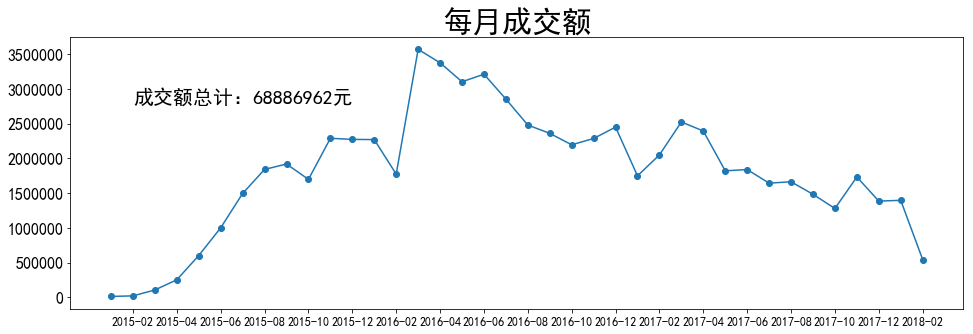
### 1.5 成交额及话题价格相关分析

对于任何产品，营收和利润都是大家最关心的话题。在行目前没有对行家和用户收取费用，因此也没有利润可言。但是在创世团队的设计中，在行未来将采用对约见费用抽成的方式获取收入。那么现在就可以看一看在行上约见的成交额及价格的相关数据，因此对在行的盈利前景进行分析。也可以对在行及所处市场的性质有更多了解。

本节分为两个部分，分别是成交额时间分布以及成交额及约见次数的价格分布。

#### 1.5.1 成交额时间分布

首先来了解一下在行上成交额的月度分布。



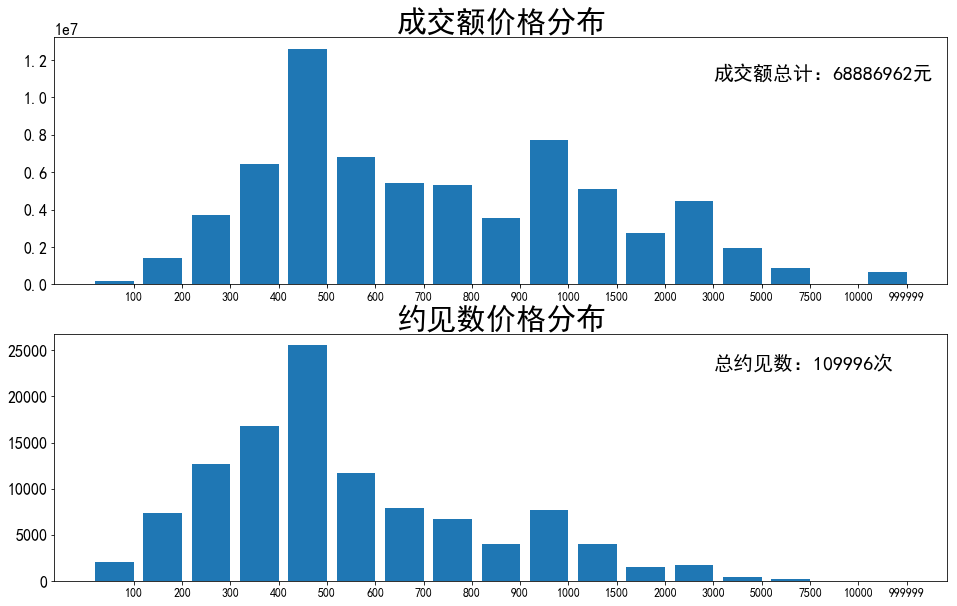
数据来源于评论信息，因此样本成交总额与总体有一定差距，但是不影响分析。

结合之前的数据，大部分月份，在行的月成交额应该在200万到300万之间。

对比这条曲线与每月评论数量曲线，两者相似的分布说明在行的次均约见价格一直保持不变。这对在行来说是好消息，因为在行的客单价本来就比较高，当务之急是在保持高客单价的条件下扩大用户规模。

#### 1.5.2 成交额及约见次数的价格分布

之前提到，在行的客单价约为615元，这是个很高的数字。这个数字真实吗？在行上话题的价格不一，作为与众不同的“知识技能共享平台”，其用户是更关心价格，还是关心约见的质量和收获？



如图横坐标为话题价格的不同区间，纵坐标为这个区间内的成交额或约见数的累计。由于大部分话题的价格在1000以内，因此1000以内的区间划分较为密集，在1000以后价格区间的范围逐渐上升，以免出现大量空白区间。

通过上图可以看到，在行上的成交额，绝大部分是由价格500及以上的话题贡献的。以400-500这个区间为分界线，右侧的累计成交额远大于左侧。而从约见数量上看基本上也能够和左侧持平。而900-1000区间的约见数甚至略高于100-200区间。而在大于10000的区间，也有20次之多的约见。如此可见，在行615的评价约见价格是有根据的。

随着价格的变化，约见数量的变化也比较平缓，可见在行的用户对于价格不是特别敏感。而400-500区间的约见数之所以最多，和那可能是因为实现了行家和约见者心理价位的平衡。

在在行上，随着话题价格的下降，约见数量并没有一直上升，当价格低于500时，价格的降低反而会导致约见数量的减少。由于这里不涉及每个区间的话题数量，因此造成这种现象的原因，一是因为行家不愿意用低价来吸引约见者，二是约见者也会担心低价的约见能否给自己有效的帮助。

在行的约见需要行家和用户付出很大的时间和精力成本，因此不是能够轻易扩大规模的产品。而从价格与成交之间的关系来看，一味降低价格，也不会带来约见数量的巨大提升。即便是约见次数最多的400-500区间，也没有比其他区间高出很多。这表示，在行上话题的定价原则应是根据话题内容和定位实现精准定价，降价促销会有帮助，但永久性的降价则可能不合适。

#### 1.5.3 总结

在行的月成交额约为每月200万到300万的水平，如果以抽成的方式来赚取收入，如果为10%那么就是20到30万，20%就是40到60万。即使不考虑抽成对成交额的影响，这样的收入依然十分薄弱。如果不能扩大约见规模，那么在行的盈利前景比较悲观。

在行上的用户并不一味追求低价，过低的价格会导致约见人数下降。总体来看在行用户对价格不敏感。在行应追求定价的精准和人性化，并适当增加定价的弹性，以便满足不同用户的需求。

### 1.6 在行产品分析总结

在行开放三年，在产品成交额和流量上做到了中等的规模，但是活跃用户的规模却偏小。产品的热度在产品出现一年之后便到达顶峰，这限制了产品规模的扩大。

在行的用户的反馈比较积极，有一定的忠诚度，但是新用户和留存用户的规模都在不断下降。

在行的高门槛和产品复杂性限制了用户群体的扩大，但在行也因此实现了较高的用户价值。不过由于规模的限制，在行的盈利前景很不乐观。

在这一章，我们根据在行的数据，分析并描述了在行的运营情况，但并没有做深入分析。因此只能得到简单的描述性结论。接下来我们会从特定方面出发，分析在行的具体细节。

注：标有（修正）的数值并非从数据中直接得出，而是经过了调整之后的估计值，同时它也表示此数值所表现的是总体的属性，而不仅是抓取到的样本的属性。

注2：对数据真实性、完整性和准确性有疑问的读者，请参阅 附录1 数据检阅及可靠性分析 的相关内容。

## 2 市场分析

在行的模式，类似于O2O与咨询行业的结合。在行把咨询的流程中，除了见面交流之外的所有环节搬到线上，并利用线上交流的便利性和资源的丰富性来扩大市场。

但在行上的约见最多只能算是简化版的咨询：流程短、跟踪反馈相对较弱、主要的交流沟通都在一次见面中完成。同时内容的提供方也分散到各行各业的众多独立个体，在大大提高了内容的丰富性的同时，也降低了内容的复杂性和深度。

需要认识到，在行的出现开辟了一个新的市场领域。就是在闲余时间，作为主业的副产品提供的个人咨询市场。在行的行家之所以吸引人，很大程度上是因为他们是某个领域的佼佼者。

新的市场，意味着新的行为特征，新的理念，新的财富和新的困难。

在研究分析在行的供给和需求之前，我们先来分析一下在行所处的市场有什么特征。看一看那些由所有参与者共同决定的要素。

本章中，我们将选择地域和分类这两个受客观因素影响的角度，通过分析在行数据内部的一些关联和区别，尝试对在行所处的市场进行分析，产生基础的了解。

### 2.1 地域分析

在行的模式是线上联系，线下约见，因此行家和约见者一般都在相同的城市中。如果想在在行上约见行家，首先要选择的就是行家所在城市。

在行可以选择入驻哪些城市，但每个城市最终带来的收益，并不完全由在行控制。这也会反过来影响在行入驻城市的数量和具体城市的选择。借此，可以了解在行所处市场的一些信息。

目前，在行开通服务的城市有：北京、上海、深圳、广州、杭州、成都、西安、武汉、宁波，共九个。

城市的数量，和在行的发展态势是相关的， 发展的越好，城市数量就越多。

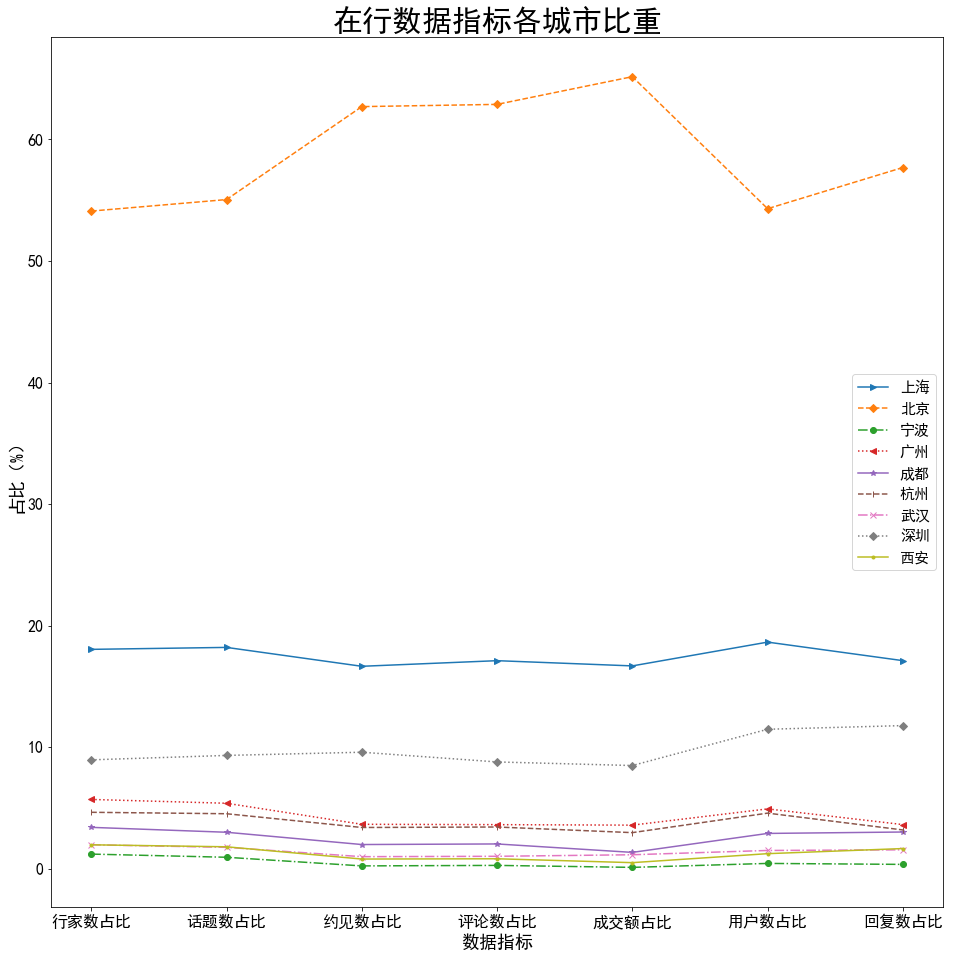
但是在行的模式又是特殊的。

具体来讲，在行家方面，经济越发达的城市，越能够提供具有丰富经验和专业技能的高端人才；观念越先进、开放的城市，高端人才愿意成为行家的比例就越高。而在约见者方面，平均收入越高的城市，有能力约见行家的人越多；发展越快速的城市，需要寻求行家帮助的问题就越多。

因此，不能简单地根据城市的数量判断在行的发展情况。而需要从各城市的数据的对比，来对在行在各城市的发展形势做出分析。

针对这种情况，我们选取了一些关键指标，以各城市的数据占这项数据总和的比例，来表示这项数据的评分。例如，假设在行行家共有1000人，而北京有200位行家，那么北京的行家数量分数为20分。这种方法的优点在于能够直观地比较：1，同一城市不同口径的数据的相对区别；2，同一数据下的不同城市的区别；以及3，不同城市的不同口径的数据的关联和区别。

详情如图所示：



可以看到：

首先，我们能够对比北京在不同指标下的评分。比如北京的行家数为54分，即占比54；约见数占比为62分。我们发现一个城市在不同指标的评分是不一致的，这是横向比较。

其次，我们能够看到一个指标下不同城市的评分，如行家数北京为54分，上海18分，宁波则只有1.2分。不同的城市存在巨大差异，这是纵向对比。

最后，北京的约见/行家的分数比为62/54=1.148,上海为16/18=0.889。那么对于约见数/行家数，即行家的人均约见数，北京是上海的1.15/0.89=1.29倍，这是交叉对比。（可以由公式推导得出，这样使用比例计算的倍率与采用真实数值计算的倍率相等。）实际上，我们也可以通过观察图中对应数据点之间连线的斜率，来比较像人均约见数这样的二级指标的大小。

由于这张图中不涉及各项数据的真实数值，因此仍是有局限性的。但是在本部分，我们主要分析数据整体内部的关联和区别，因此这种分析方法在此是适用的。

仔细分析上图之后，我们可以得出两个主要结论：

1．在行在不同城市的规模是十分不均衡的。北京一个城市，为在行做出了50%-60%综合贡献。但九个城市中的六个，综合贡献不足5%。

参考每个城市首个评论出现的日期：

|  |  |
| --- | --- |
| 北京 | 2015年1月 |
| 上海 | 2015年2月 |
| 杭州 | 2015年3月 |
| 深圳 | 2015年4月 |
| 广州 | 2015年4月 |
| 成都 | 2015年5月 |
| 宁波 | 2015年8月 |
| 武汉 | 2015年9月 |
| 西安 | 2016年3月 |

在行进驻每个城市的时间长短的区别，远远不足以解释各个城市之间的巨大差异。因为我们选择的指标涵盖了约见的供给方和需求方，我们可以说，这里展现出的差异，就是在行所处的市场在不同城市的差异。

作为参考，2017年中国百强城市排行榜中各城市的排名为：

北京：1、上海：2、深圳：4、广州：3、杭州：7、成都：6、西安：16、武汉：10、宁波：20。

文章链接：<http://henan.163.com/17/1120/15/D3MPL1G604398DOB.html>

2．在行上，北京的活跃度高于其他城市，而除了北京和深圳外，其余城市的活跃度比较相似。

这里的活跃度指约见数、评论数、成交额和回复数四个指标，它们大多与用户的积极程度有着关联。与之相对的则是表示数量指标的行家数、话题数、用户数三个指标。

影响活跃度的主要是用户群体的认可度和满意度；而影响数量指标的，不仅有在行的决策和投入，更有每个城市的硬性条件，如经济发展水平，行业成熟度等。

可以看到，北京的活跃度评分是要高于数量评分的，其余城市则大多相反。

根据这两个结论，需要思考，为什么在行在北京以外的城市推进不顺利？为什么北京会展现出与其他城市不同的特征？

原因主要在于城市条件的区别，但也与在行的运营有关。

首先因当考虑在行的模式是否不够成熟，不够标准化和普世化，以至于没能走出北京这个全国的互联网中心。其次，在行在其他城市的运营是否有不妥当的地方，导致无法完全挖掘出市场的潜力。最后，是否是城市当地的一些因素，影响着市场的成熟度，以及在行的发展。

我们看到，在行的表现，以及在行所处的市场，在不同城市确实是有差别的。

针对城市间的区别，还有许多可挖掘的地方。考虑到篇幅，为了避免啰嗦和繁杂，本节先就分析到这里。

最后附上上文例图的具体数据：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **行家数** | **行家数占比** | **话题数** | **话题数占比** | **约见数** | **约见数占比** | **评论数** | **评论数占比** |
| **上海** | 1485 | 18.048128 | 2725 | 18.209155 | 24557 | 16.655024 | 18828 | 17.116986 |
| **北京** | 4452 | 54.107924 | 8239 | 55.055129 | 92452 | 62.702703 | 69173 | 62.886832 |
| **宁波** | 99 | 1.203209 | 141 | 0.942198 | 342 | 0.231951 | 301 | 0.273646 |
| **广州** | 469 | 5.700049 | 806 | 5.3859 | 5390 | 3.6556 | 3989 | 3.626496 |
| **成都** | 280 | 3.403014 | 449 | 3.000334 | 2934 | 1.989895 | 2242 | 2.038256 |
| **杭州** | 382 | 4.642684 | 677 | 4.523889 | 5003 | 3.39313 | 3775 | 3.431943 |
| **武汉** | 162 | 1.968887 | 263 | 1.757434 | 1462 | 0.991556 | 1125 | 1.022764 |
| **深圳** | 737 | 8.957219 | 1395 | 9.321751 | 14132 | 9.584591 | 9663 | 8.784865 |
| **西安** | 162 | 1.968887 | 270 | 1.80421 | 1173 | 0.795551 | 900 | 0.818212 |
| **总和** | 8228 | 100 | 14965 | 100 | 147445 | 100 | 109996 | 100 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **成交额** | **成交额占比** | **用户数** | **用户数占比** | **回复数** | **回复数占比** |
| **上海** | 15143324 | 16.686791 | 10684 | 18.644097 | 4229 | 17.119378 |
| **北京** | 59137891 | 65.165455 | 31119 | 54.304162 | 14252 | 57.693398 |
| **宁波** | 100514 | 0.110759 | 250 | 0.436262 | 88 | 0.356232 |
| **广州** | 3253539 | 3.585152 | 2821 | 4.922782 | 897 | 3.631138 |
| **成都** | 1219808 | 1.344136 | 1664 | 2.903761 | 746 | 3.019876 |
| **杭州** | 2694400 | 2.969024 | 2621 | 4.573772 | 788 | 3.189896 |
| **武汉** | 1043125 | 1.149444 | 861 | 1.502487 | 384 | 1.554467 |
| **深圳** | 7704547 | 8.489824 | 6575 | 11.473693 | 2909 | 11.775898 |
| **西安** | 453221 | 0.499415 | 710 | 1.238984 | 410 | 1.659717 |
| **总和** | 90750369 | 100 | 57305 | 100 | 24703 | 100 |

### 2.2 行业分析

除了以地区区分之外，在行还对话题进行内容的分类。分类标准包括行业、领域、职能、话题类型等等，比较复杂，但总体上是按照话题内容或功能的不同而区分的。

在行的分类分为3级，分类关系为：一级分类包含二级分类，二级分类包含三级分类。

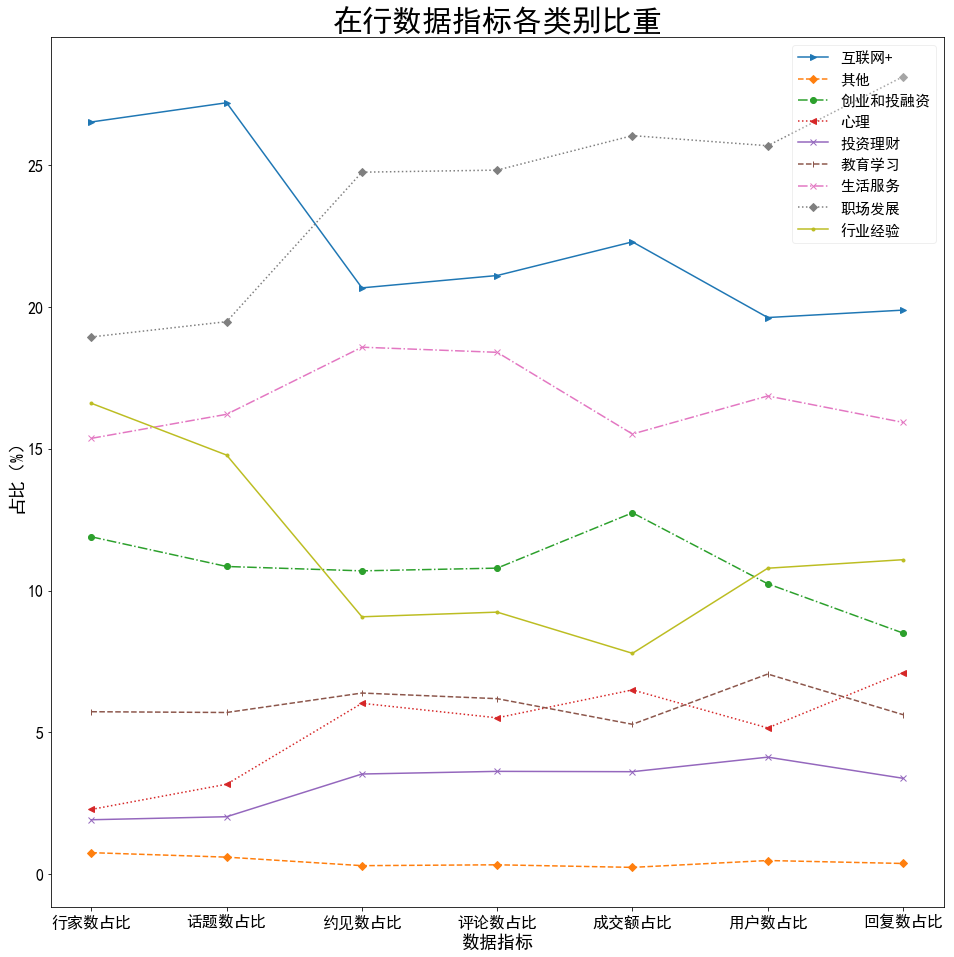
一级分类有9个，分别是：职场发展、行业经验、互联网+、创业和投融资、生活服务、心理、投资理财、教育学习、其他。二级分类有82个。三级分类86个。总计177个。

出现在上级分类的话题不一定出现在下级话题，反之则不同。同时一部分二级话题并没有继续细分为三级话题。

内容的分类，便于在行进行精准运营和管理，也便于约见者挑选自己心仪的内容。但分类做不好，则可能会产生负面影响。

话题分类的情况，以及各分类的表现，一方面受在行控制，另一方面也受用户需求和行业自身发展的影响。按照本章序言的标准，我们可以将话题分类的情况看作是受市场因素的影响。

在本节我们使用的方法与上节相同，区别是将城市换成了话题分类。同时，由于一个话题或行家可能在不同分类出现，统计总量会大于话题或行家的总数量。为了分析方便，本节至对一级分类进行分析。



图表内容和使用方法的解释不再赘述，请参见上节。

观察这幅图，得到的结论有：

1. 各分类的规模较为均衡。应该是人为所致，因为若是规模太小，便不足以成为一级分类。
2. 各分类在活跃度和数量指标的相对表现各不相同。这点在职场发展和互联网+这两个分类上尤为明显，职场发展所占的话题数和行家数都少于互联网+，但其余指标却全部反超。根据指标的相对变化，可以说，活跃度大于数量指标，说明需求相对旺盛，供给不足；反之，供给相对旺盛，需求不足。市场的真实需求，与在行的规划并不完全吻合。
3. 各分类的表现也有参考价值。排名靠前的职场发展、生活服务、互联网+、创业和投融资，有这样几个共同特征：跨领域、内容回报周期短、内容回报价值大。这说明用户喜欢短周期、价值明确的内容，市场的成熟度还很低。而行业经验低于预期的表现，很可能是因为与工作相关的专业问题，很多情况下都可以通过业内或公司的资源解决。这点可能是许多人没有料到的。

同样附上数据：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **行家数** | **行家数占比** | **话题数** | **话题数占比** | **约见数** | **约见数占比** | **评论数** | **评论数占比** |
| **互联网+** | 2834 | 26.5182 | 3989 | 27.195255 | 30110 | 20.669156 | 22897 | 21.104782 |
| **其他** | 81 | 0.75793 | 88 | 0.599945 | 437 | 0.299981 | 358 | 0.329978 |
| **创业和投融资** | 1271 | 11.892954 | 1591 | 10.846741 | 15577 | 10.692908 | 11705 | 10.788814 |
| **心理** | 244 | 2.283148 | 465 | 3.170166 | 8775 | 6.023642 | 5979 | 5.511005 |
| **投资理财** | 205 | 1.918218 | 297 | 2.024816 | 5143 | 3.530437 | 3932 | 3.62423 |
| **教育学习** | 612 | 5.726584 | 836 | 5.699482 | 9302 | 6.385403 | 6712 | 6.186631 |
| **生活服务** | 1642 | 15.364461 | 2378 | 16.212163 | 27063 | 18.577528 | 19959 | 18.396748 |
| **职场发展** | 2024 | 18.938898 | 2857 | 19.477775 | 36051 | 24.747385 | 26928 | 24.820263 |
| **行业经验** | 1774 | 16.599607 | 2167 | 14.773657 | 13218 | 9.073561 | 10022 | 9.237547 |
| **总和** | 10687 | 100 | 14668 | 100 | 145676 | 100 | 108492 | 100 |
|  |  |  |  |  |  |  |  |  |
|  | **成交额** | **成交额占比** | **用户数** | **用户数占比** | **回复数** | **回复数占比** |  |  |
| **互联网+** | 20001038 | 22.291243 | 14080 | 19.623146 | 4859 | 19.884596 |  |  |
| **其他** | 214669 | 0.23925 | 345 | 0.480823 | 92 | 0.376494 |  |  |
| **创业和投融资** | 11430725 | 12.739592 | 7345 | 10.236648 | 2077 | 8.499754 |  |  |
| **心理** | 5828315 | 6.495682 | 3699 | 5.155257 | 1737 | 7.108365 |  |  |
| **投资理财** | 3241271 | 3.612411 | 2960 | 4.125321 | 826 | 3.380259 |  |  |
| **教育学习** | 4740693 | 5.283523 | 5062 | 7.054856 | 1374 | 5.622852 |  |  |
| **生活服务** | 13922185 | 15.516335 | 12096 | 16.858067 | 3892 | 15.92732 |  |  |
| **职场发展** | 23361329 | 26.036302 | 18426 | 25.68012 | 6870 | 28.114258 |  |  |
| **行业经验** | 6985762 | 7.785662 | 7739 | 10.785762 | 2709 | 11.086102 |  |  |
| **总和** | 89725987 | 100 | 71752 | 100 | 24436 | 100 |  |  |

### 2.3 附言

本章的数据还有很多待研究和挖掘之处，包括对分类和城市进行交叉分析，以及对分类和城市分别进行时间序列分析。但这些分析的复杂度会大大增加，适合作为专题分析报告单独出现。在此暂不进行深入研究。

### 2.4 总结

在行的发展受到市场因素的影响，具体表现为城市之间市场成熟度的不同，从而导致在行在不同城市的发展不均衡。在行的话题分类的数据展示出在分类上可能存在供需的错配，这可能是在行发展受阻甚至衰退的原因之一。

## 3 供给分析

附录1 数据检阅及可靠性分析

此次采用的数据并非百分百完整和准确，为了使数据分析的前提得以成立，首先需要对到手的数据做一些分析和研判。

首先来验证一下数据的完整性。

分别在百度和谷歌搜索‘site:www.zaih.com/mentor’,这表示在在行的域名中寻找行家页面。

百度搜索结果约6300条。



谷歌搜索结果约19300条。

但是谷歌的搜索结果中同时包含了行家页面和话题页面。





此次抓取的行家和话题的总数为13204条，大于19300。因此本报告采用的数据涵盖面要大于百度和谷歌的收录范围，再根据数据之间的互相印证，数据的完整性应当达到了90%-97%，足以支撑一篇具有参考价值的数据分析报告。