



# 数据分析

CHINA DATA ANALYSIS 数据分析·因你而不凡



《中国数据分析》行业特刊  
2020年第04期 总第44期 (季刊)  
咨询热线: 400-050-6600  
<http://www.chinacpda.org/>  
投稿邮箱至 xiehui@chinacpda.org



China Data Analysis

## “数”有所为 赋能新时代 点燃新引擎

从2015年党的十八届五中全会提出“实施国家大数据战略”开始，持续性的方向引导和顶层设计，使我国在大数据发展规划布局、政策支持、资金投入、技术研发、创新创业等方面均走在了世界前列。这些都表明，我国运用大数据推动创新发展具备较好的基础和条件，也说明党和国家实施大数据战略的重要性和正确性。

2020年，以新一代信息技术为基础的数字经济已经成为我国经济发展中最为活跃的领域，是我国国民经济高质量发展的新动能。在前不久，党的十九届五中全会上，针对“加快数字化发展”作出全面部署，要求推进数字产业化和产业数字化，推动数字经济与实体经济各领域的深度融合所带来的生产效率的提升以及生产模式的改变，成为产业转型升级的重要驱动力。积极抓住数字产业快速发展的机遇，推动大数据与传统行业的深度融合，利用数据资源分析，研判企业发展趋势，辅助政府精准决策，支撑行业转型升级，构建起基于互利共赢的产业新生态。

今年年初，新冠肺炎疫情突袭。习近平总书记在疫情防控工作时表示：“要运用大数据等手段，加强疫情溯源和监测。”疫情实时大数据报告、追踪疾病接触人员动态、智能调度医疗防护资源……疫情让数字产业抓住了机遇，传统产业也被推动着向数字化转型、智能化升级。受新冠肺炎疫情影响，国际经贸活动受到严重影响，我国经济发展面临新挑战，面对复杂的经济形势，危中寻机、化危为机，抓住产业数字化、数字产业化赋予的机遇，大力推进科技创新。后疫情时代，中国经济构建出新的发展格局。

虽然我国大数据快速发展的格局基本形成，但在数据开放共享、核心技术突破、以大数据驱动发展等方面都面临重重挑战。发展之路没有终点，只有新起点。大数据时代，数据正在成为一种生产资料。任何一个行业和领域都会产生有价值的数，而对这些数据的统计、分析、挖掘则会创造更多的价值和财富。在海量数据背后，是巨大的信息价值和洞察力。数据汇聚使数据可能产生价值，数据分析使数据实现价值，完成从数据到信息再到知识和决策的转换，才能有力推动经济转型。未来数据分析的发展，将不可限量！

2020年注定会在历史长河中烙下印记，大数据的强大赋能，数据分析的决策引导，正前所未有的促进各行各业价值创造方式。作为行业协会，始终不忘初心，砥砺前行，以促进国家大数据战略目标实现为总原则，在规范从业行为、倡导行业自律、培养大数据人才，引领行业不断向前可持续性健康发展的道路上，不断审时度势，开拓创新、劈波斩浪、行稳致远。

心有所信，方能行远。2021年即将到来，在坚守中与时俱进，在“数”有所为的智能时代，我们共同守正笃实，久久为功。

站上新起点，扬帆新征程，赋能新时代，点燃新引擎！

中国商业联合会数据分析专业委员会

## 本期目录 CONTENTS



### 主办单位

中国商业联合会数据分析专业委员会

### 编委成员

李苗苗

### 出版时间

2020年12月出版 总第44期

### 美工设计

崔峻珩

### 联系我们

中国商业联合会数据分析专业委员会  
地址: 北京市朝阳区朝外SOHO-C座9层  
电话: 400-050-6600 / 010-59000991  
传真: 010-59000991转 607

### 卷首语

- 01 "数"有所为 赋能新时代 点燃新引擎

### 协会动态

- 03 CPDA数据分析师|请查收你的高光时刻!  
05 强强联手共同推进金融行业数据分析人才专业化进程  
06 9月11日 CPDA沙龙回顾 | 数据分析如何在零售时尚行业落地  
08 9月25日 CPDA沙龙回顾 | 如何做好投资分析与评价  
10 10月22日 CPDA沙龙回顾 | 如何做好销售预测解决方案  
11 11月12日 CPDA沙龙回顾 | 如何通过推荐系统实现业务增长

### 政策向导

- 13 中国科学院全力建设可持续发展大数据国际研究中心  
14 国企数字化转型驶上快车道: 组建国资大数据产业集团设立大数据产业基金  
15 《北京国际大数据交易所设立工作实施方案》等重磅政策发布  
16 《贵州省政府数据共享开放条例》12月1日起施行  
20 深圳综合改革试点实施方案发布支持建设粤港澳大湾区数据平台

### 行业动态

- 23 让"大数据+医疗"更好满足群众需求  
24 大数据风控将助推网贷中介平台转型  
25 善用大数据让老赖没法赖  
26 大数据技术为制造业提供多方位、精细化的服务  
27 治理"大数据杀熟"重在对症下药

### 学"数"交流

- 29 深度抽吸对卷烟烟气指标影响的研究  
33 Python提速的10个小技巧, 你值得拥有  
38 阿里巴巴大牛是怎么写数据分析报告的?  
41 数据分析实战案例: 渠道落地页效果分析  
45 20种数据可视化工具测评, 有一半你都不知道  
49 Excel做"回归分析", 你可能都没玩儿过!

### 事务所专栏

- 53 CPDA学员专访 | 周召安: 成就源于平凡的选择和不放弃的努力  
56 山东智谷数据分析师事务所

欢迎广大读者踊跃投稿, 内容包括学术观点、教学体验、教学活动、学习感悟、实战经验、随笔文章等。

稿件附图格式为JPG或TIFF格式, 大于1M, 分辨率在300dpi以上。

感谢您对《中国数据分析》的支持! 投稿邮箱: xiehui@chinacpda.org

# / CPDA数据分析师|请查收你的高光时刻! /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-10



近期，我们收到了很多学员的建议，大家纷纷表示目前的证书更侧重功能性，在年检和入会注册方面很方便，但欠缺展示效果。希望《CPDA数据分析师证书》可以与国际接轨采用电子版证书：

“我要把证书摆在我的工作台上，每天激励自己。”“在外企工作，证书一定要国际化起来，现有证书不太方便展示。”“客户看到我办公室墙上挂着证书，一定会对我们更有信心。”“我的高光时刻必须要在朋友圈秀一把，收获一大波崇拜。”

听到这些心声，协会决定华丽上线CPDA数据分析师电子证书，证书采用中英文双语，设计精美，更突显个性化和仪式感。为了更好的服务广大数据分析师，电子版证书显示了有效期，便于学员及时年检享受各类专属福利。证书有效期内的学员可以免费参加全国性行业峰会、免费参加每月1-2次的专属沙龙活动，免费申请复听课程，也可以免费使用智能大数据分析平台——Datahoop。

电子版证书作为纸质版证书的补充，而非替代品，与纸质证书具有同等效力。今后凡是通过CPDA数据分析师考试的

学员，都将获得纸质版+电子版双证书。



图：《CPDA数据分析师证书》电子版样式

电子版证书采用国际化标准设计，体现中英文双语内容，兼有中国商业联合会数据分析专业委员会红章和CDAC电子签章认证。学员可自行下载、打印装裱陈列在办公空间、或陈列展示空间。《CPDA数据分析师证书》是对个人专业能力的认可，它彰显着“数据分析师”的价值和过硬的执业水准。

它不仅是个人职业生涯中里程碑式的奖章，更是数据分析师鞭策自己不断前行的动力。



图：《CPDA数据分析师证书》打印效果展示

CPDA数据分析师纸质证书将继续保留方便学员、会员进行年检。

CPDA数据分析师电子证书华丽上线，并不会代替纸质证书。纸质证书仍将继续承担着学员的CPDA证书年检以及协会会员会籍管理两大职能。

### 证书年检

CPDA数据分析师证书需要每三年进行一次年检，年检完成后会在年检页加盖“年检贴”。年检更新后，学员可同步登录CPDA数据分析师官网下载最新有效期的CPDA数据分析师电子证书。



图：《CPDA数据分析师证书》年检页

### 会籍管理

CPDA数据分析师一经加入会员，需要在纸质证书上完成个人注册信息的登记和变更，包括每年一次的会员年检。



图：纸质证书会员注册信息页 样式

### 行业权威性及市场认可度

《CPDA数据分析师证书》是中国商业联合会数据分析专业委员会颁发的数据分析行业执业证书，同时具备创办事务所、政府人才引进、公务员报考、招投标加分的价值及影响力。

- △优势1：享有多地公务员考试加分待遇
- △优势2：政府、企业数字化项目招投标必备
- △优势3：持证者优先入选《中国大数据人才库》
- △优势4：申请成立数据分析师事务所必备证书

《CPDA数据分析师证书》的行业权威性与市场认可度，源于每一位CPDA数据分析师的共同努力。你们在各自岗位上专注大数据的实战能力、致力于挖掘大数据的实际应用价值，用大数据思维助力企业数字化转型。是你们推动着中国数据分析更快、更强的发展！

# / 强强联手共同推进金融行业数据分析人才专业化进程 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-12



2020年12月3日，中国商业联合会数据分析专业委员会督导下的高端数据分析人才认证培训项目‘CPDA数据分析师’与北京民生财富研修学院达成深度合作，强强联合共同推进我国金融行业数据分析人才专业化进程。

课程产品的认可，北京民生财富研修学院将成为CPDA数据分析师的又一合作渠道，结合双方各自优势和资源，构建金融行业全国范围数据分析师认证人才的“集训大本营”，培养具备数据认知能力、数据处理能力、数据化思维能力、数据呈现能力、数据决策能力、计算机及数据分析信息技术和企业实战能力的优质数据分析师专业人才！



中国商业联合会数据分析专业委员会 会长 邹东生先生  
与北京民生财富研修学院 院长 王晓刚先生  
签订合作协议



CPDA数据分析师是中国数据分析类人才认证第一品牌。自2003年起，已为各行业输送了数万名数据分析师。遍布在各行业的数据分析师们运用大数据思维，为政府和企事业单位在经营和决策过程中提供了科学完整的数字化解决方案，成为驱动各行业数据落地应用的先行者。

### 背景介绍：

北京民生财富研修学院是由中国民生银行发起，北京民生文化艺术基金会及多家股份制企业共同出资创办，学院为教育领域非营利性组织，致力于金融行业人才教育公益事业。依托先进的办学模式、借助国内外顶尖教育资源，开设全球金融EDP项目、全球金融专家项目等高端培养项目，致力于培养高层次、创新型、国际化的金融人才；同时开设专业课程类、特色创新项目、金融资质类项目及金融论坛等，为金融机构提供强有力的智力输出与实战指导。

2020年初，CPDA数据分析师走进民生银行，通过专业的师资讲授、真实的案例讲解、结合大量的实操演练顺利完成了300余人的认证培训工作，取得了广大学员的一致褒奖，纷纷表示课程内容极具实战指导性。缘于对人才培养理念及专业化

# / 9月11日 CPDA数据分析师沙龙回顾 数据分析如何在零售时尚行业落地 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-09



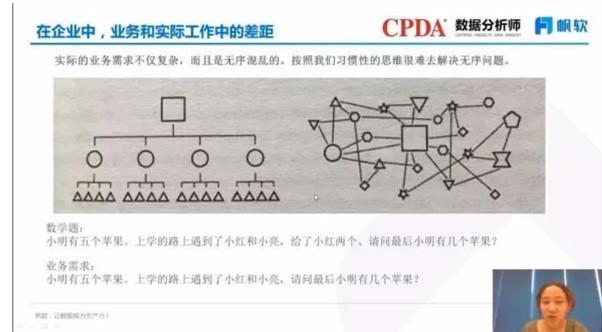
9月11日，一场主题为《数据分析如何在零售时尚行业落地》的CPDA专属沙龙活动如约举行。在数百名的沙龙参与者中，既有身处零售&电商领域的数据分析师，也有从事相关课题研究的从业者，他们与CPDA特邀嘉宾马天睿老师共聚沙龙直播间，共同探讨交流，展开了一场以提高零售业务为核心的实战演习。

零售业是感受市场风向最敏锐的行业之一，在进行数字化转型的过程中，面临的问题更加具体也更加棘手。为了让活动更具有工作指导意义，沙龙内容聚焦零售业最实际的数据分析工作，以多个具体实践案例，帮助大家切实掌握从初级需求梳理到决策支撑，再到落地执行的全套工作流程。不少同学在活动后表示，本期活动的实操案例正对自己的业务要求，解决了自己一贯以来的工作困惑，让自己茅塞顿开。

在一个多小时的沙龙活动中，整体气氛活跃，答疑环节讨论热烈，大家收获满载。现场还有哪些精彩瞬间，让我们一起来回顾吧！

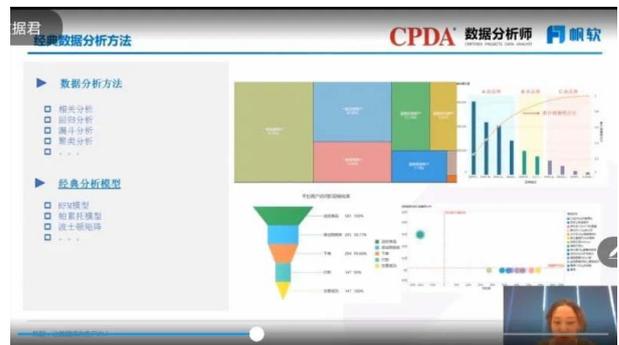
## 01. 如何贴合业务进行数据分析

首先，在分享内容的第一部分，马老师先抛出了“如何贴合业务进行数据分析”的问题，由此为大家引出了分析业务的一串流程步骤，包括明确分析需求、制定分析计划、数据拆分模型、产出商业决策、验证决策效果等等。进而讲到管理业务需求、制定分析计划，以及趋势分析法、多维分析法、经典分析模型等重要内容。



## 02. 超市经营问题诊断分析案例

工具理论铺垫之后，马老师开始以上海徐家汇附近一中型超市的经营问题为例，通过排除非核心要素，带领大家从三个步骤综合考虑购物篮系数的问题。首先考量的是购物篮系数与时间的关系，其次是购物篮系数与顾客购买行为的关系，第三是购物篮系数与商品缺货的关系，层层递进，最终给出方案及成果。





## / 9月25日 CPDA数据分析师沙龙回顾 /

### 如何做好投资分析与评价 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-09



9月25日，协会特邀中国注册会计师、经济师王德海老师为我们的学员带来一场《如何做好投资分析与评价》主题沙龙分享。

在一个多小时的沙龙活动中，王老师不仅从盈利能力分析、偿债能力分析、财务生存能力分析、不确定性分析等众多角度为学员们详细讲解了财务分析知识框架，还通过实际案例对编制报表、项目投资估算流程、总成本估算方法等方面的知识应用展开了深入的分享。

王老师的分享深入浅出，干货满满。数据君将沙龙中的知识点整理成了学习笔记，分享给大家。

#### 知识点1 掌握公司估值模型DCF

对于创业公司，往往需要到资本市场上寻求投资，这就涉及到公司估值的问题，而公司估值最基本的方法就是折现现金流模型，英文简称DCF。财务投资分析同样需要用到DCF模型，你掌握了相关的知识就能看懂公司估值，进而能判断出估值是否合理。

$$P = \sum_{t=1}^n \frac{CF_t}{(1+r)^t}$$

- P—企业的评估值；
- n—资产（企业）的寿命；
- CF<sub>t</sub>—资产（企业）在t时刻产生的现金流；
- r—反映预期现金流的折现率。

CF模型的基本原理：金融学里认为，资产之所以有价值，在于它未来能给我们带来一系列现金流入，将这些现金流入折现、加总，就能得到这些资产当前的价值。

#### 知识点2 4个维度做好财务分析

在做项目投资时，一般可以从四个维度来做财务分析，分别是：盈利能力分析、偿债能力分析、财务生存能力分析、不确定性分析，每一项分析都有相应的指标，并且涉及多个数据报表。

#### 从哪几个角度对项目进行财务分析？

##### 一、盈利能力分析

(1) 涉及报表：项目投资（全投资）现金流量表、资本金现金流量表、投资方现金流量表、利润与利润分配表

(2) 评价指标：净现值、投资回收期、内部收益率、投资利润率等

##### 二、偿债能力分析

(1) 涉及报表：资产负债表、还本付息计划表。

(2) 评价指标：资产负债率、流动比率、速动比率、偿债备付率、利息备付率等

##### 三、财务生存能力分析

(1) 涉及报表：财务计划现金流量表

(2) 评价指标：累计盈余资金

##### 四、不确定性分析：敏感性分析、盈亏平衡分析、概论分析

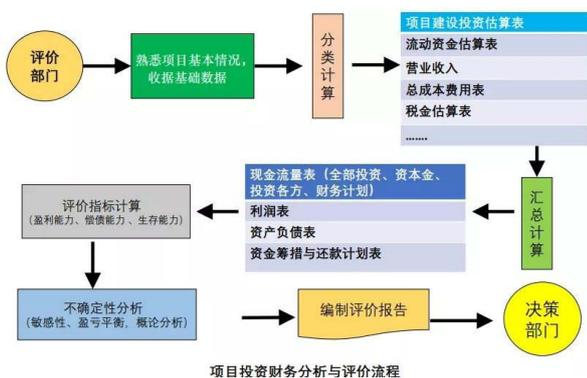
· 盈利能力分析涉及各种现金流量表和利润分配表，它的评价指标是净现值、投资回收期、内部收益率、投资利润。

· 偿债能力分析涉及资产负债表、还本付息计划表，评价指标是资产负债率、流动比率、速动比率、偿债备付率、利息备付率。

· 财务生存能力分析涉及财务计划现金流量表，评价指标是累计盈余资金。

· 不确定性分析包括敏感性分析、盈亏平衡分析、概率分析。

投资分析与财务评价的基本流程：评价部门首先要熟悉项目基本情况、收集基础数据，通过分类计算得到一系列相关辅助表格，将这些表格进行汇总计算得到一套基本报表，通过这些基本表计算出评价指标，然后进行不确定性分析，最后编写完整的评价报告提交决策部门。



### 知识点3 计算假设与参数选取因素

1) 估算价格的选择一种是固定价格，另一种是变动价格。如何区分这两种价格呢？

固定价格&变动价格：如果在建设期的投资估算中，看到“涨价预备费”，说明项目采用了变动价格，如果没有这一项就是固定价格；在项目运营期，两种价格都可以使用。

含税价格&除税价格：营改增之后，建设期这些进项税是可以扣除的，进项税的留抵和以前不一样了；所有的价格都要做到“价税分离”，便于计算和分析。

2) 关于项目计算期的选取包括建设期和运营期，运营期又分投产期和达产期。运营期的计算主要是根据产品和主要设备的寿命来确定，一般不超过20年；如果寿命期很长，一般是把计算期之后的现金流折算到计算期末，再从计算期末往前折算。

### 计算假设与参数选取

#### 5、项目计算期选取

- ◆计算期包括建设期、生产运营期（投产期+达产期）。
- ◆生产运营期，应根据产品寿命期(矿产资源项目的设计开采年限)、主要设施和设备的使用寿命、主要技术的寿命期等因素确定。财务评价的计算期一般不超过20年。
- ◆项目运营寿命很长时，将计算期结束以后年份的现金流入和现金流出折现至计算期末。

3) 折现率的选择折现率本质上是一种投资报酬率，实践中常用“公式法”、“累加法”、“行业平均资金利润法”、“投资者要求的回报率”来计算（详细计算方法见下图）。注意一点：折现率不用考虑通货膨胀率，现金流要与折现率保持一致性。

### 计算假设与参数选取

#### 9、折现率的选择

折现率本质上是一种投资报酬率。实践中，确定折现率通常有以下几种方法：

##### (1) 公式法

- ◆WACC（加权平均资金成本）

$$WACC = K_e \times E/V + K_d \times (1-T) \times D/V$$

式中：K<sub>e</sub> 为权益资本成本，K<sub>d</sub> 为债务资本成本；E：权益的市场价值；D：付息债务的市值 T：企业的所得税率 V：被评估企业的总市值。

- ◆CAMP（资产资本定价模型）

$$R_i = R_f + \beta_i(R_m - R_f)$$

式中，R<sub>i</sub>为资产i的期望报酬率，R<sub>m</sub>为市场平均收益率，R<sub>f</sub>为无风险报酬率，β<sub>i</sub>为资产i的系统性风险系数，R<sub>i</sub>为资产i的风险报酬率。

#### (2) 累加法

折现率=无风险报酬率+风险报酬率

风险报酬率=行业风险报酬率+经营风险报酬率+财务风险报酬率+其他风险报酬率

#### (3) 行业平均资金利润法

如果有行业发布的本行业基准收益率，即以其作为项目的基准收益率；如果没有行业规定，则由项目评价人员设定。

#### (4) 投资者要求的最低回报率

问：折现率是否考虑通货膨胀率？

现金流与折现率保持一致性  $\frac{CF_t}{(1+r)^t}$

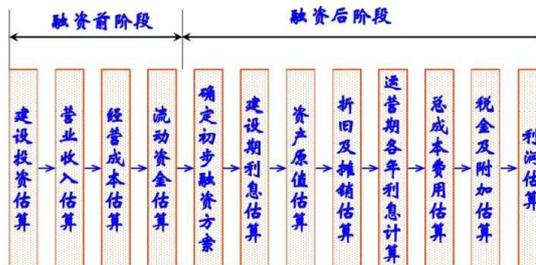
### 知识点4 如何编制报表

需要编制的主要报表，包括基本报表和辅助报表，一共有20多个。

### 需要编制的主要报表

项目投资财务分析与评价报表		
基本报表(一级表)	辅助报表(二级表)	辅助报表(三级表)
技术经济指标汇总表	项目总投资估算表	建筑工程费用估算表
全投资现金流量表(项目投资现金流量表)	流动资金估算表	建设投资估算表(形成资产法)
资本金现金流量表	营业收入估算表	设备及器具购置费用估算表
投资各方现金流量表	总成本费用估算表	安装工程费用估算表
财务计划现金流量表	税金估算表	工程建设其他费用估算表
利润及利润分配表		外购原材料费用估算表
资产负债表		外购燃料动力费用估算表
资金计划与筹措表		职工薪酬及福利费估算表
借款偿还计划表		固定资产折旧费估算表
		无形资产及递延资产摊销费估算表

首先我们要搞清楚各个报表之间的逻辑关系和勾稽关系，其次要清楚他们之间的先后次序，然后从最底层报表一级一级向上汇总，得到基本报表。



### 报表编制主线

编制报表时要重点把握两点：一是分清融资前和融资后这两个阶段，二是搞清楚建设期的支出和运营期的支出，即项目支出和成本费用。两个阶段的报表可参照下图的主线进行编制。

# / 10月22日 CPDA数据分析师沙龙回顾 /

## 如何做好销售预测解决方案 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-10



销售预测是指根据历史的销售情况以及内外部因素使用系统内部或用户自定义的销售预测模型获得的对未来销售情况的预测。企业内部因素包括：企业的营销策略、销售政策、销售人员、生产状况；企业外部因素则包括：客户的需求动向、国家经济状况、竞争对手的状况、政府的政策导向。

10月22日，协会特邀永洪高级数据分析师李洋老师为我们的学员带来《如何做好销售预测解决方案》主题沙龙分享。本次沙龙活动中，李洋老师详细介绍了“企业如何对自己的销售状况进行预测”以及“企业销售预测解决方案如何落地”。在长达一个小时的沙龙分享中，李老师给大家带来满满的干货，数据君受益匪浅，今天就把昨晚记在小本本上的知识点拿出来分享给大家。

### 知识点1：如何提高需求预测的准确性

在做销售运营计划的时候会遇到很多问题，比如各个区域的需求量和代理商实际的销售情况存在一定偏差，这种偏差会带来一定的缺货或是库存的积压。



解决这个问题需要从两方面着手：一方面是提高我们对需求预测的准确率，另一方面是我们要优化企业自身的库存。

提高需求预测的准确率需要从三方面入手：

1. 需求的感知首先拿到历史数据和其他行为数据，然后去预测建立模型、预测需求，尽可能是预测需求与市场需求匹配。

2. 需求的塑造企业内部与外部对于销售政策、价格、促销活动、经济状况等诸多因素会对实际销量产生很大影响；建立模型需要考虑这些因素，使销量、库存、生产三者相匹配。

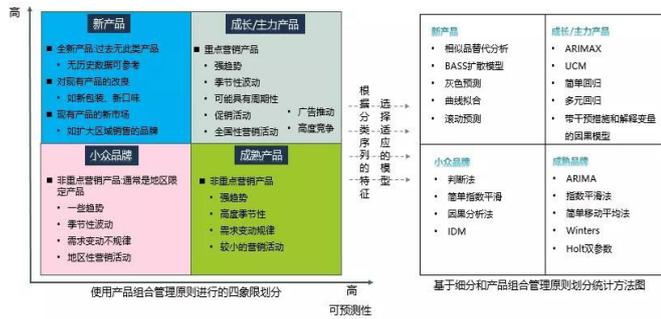
3. 需求的转移当出现缺货时，需要新产品来转移客户的需求，企业应引导客户关注新产品；库存是直接面向市场波动而设立的，当不能很好满足市场波动时，会造成库存的积压或缺货，因此库存的好坏直接影响我们对需求的满足度，需求的误差则需要通过库存来弥补。

### 知识点2：算法的选择

我们需要先将产品分为新产品、成长期产品、小众产品、成熟产品，再根据不同产品类型进行相应的算法选择。新产品由于没有历史数据，无论采用什么方法都没办法准确预测，我们可以找到和它相似的产品，基于该产品的数据进行预测，常用方法包括灰色预测、曲线拟合、滚动预测等。

成长期产品具有明显的趋势性和季节性，可以采用

ARIMAX、UCM及一些回归模型、因果模型进行预测。



小众产品由于销量小，相对受波动影响也比较小，它的可预测性也是很小的，可以用简单的判断法和因果关系法进行建模和预测。

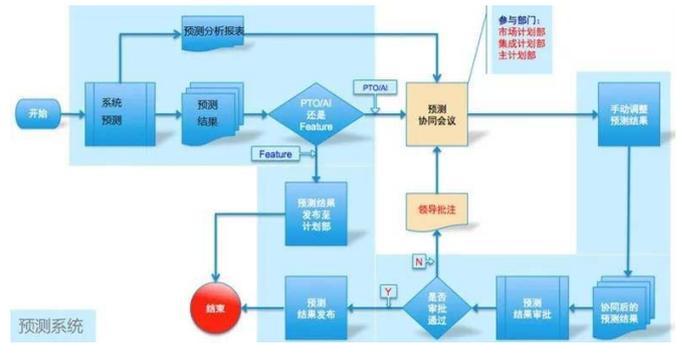
成熟期产品具有强烈的趋势性和高季节性，我们采用机器学习法和时间序列法都可以进行预测。

在做预测的时候，我们需要遵循“抓大放小”的原则，重点分析那些销量大、价值高的产品，如：新产品、成长期产

品、成熟期产品，而像“小众产品”这种可预测性低、价值低的产品可以不做预测。

### 知识点3：预测的协同流程

这个流程主要是我们对生成的预测结果需要考虑人为的因素，并对结果进行人工干预；通过协同的会议发布结果，在会议上需要记录来自市场、销售、采购几个部门对于结果干预的状况，并进行标注。后期，我们可以再调整这个人对这个结果干预的影响性和权重，使得最终的预测结果更符合实际。



## / 11月12日 CPDA数据分析师专属沙龙回顾 / 如何通过推荐系统实现业务增长 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-11

在这个数据爆炸的时代，用户对个性化服务的要求越来越高，已经不再满足于统一的用户体验，而是希望系统能够根据自己的兴趣喜好，提供相应的推荐内容以及有针对性的服务，推荐系统便是在这样的背景下应运而生。

如何利用数据分析技能在庞大的信息库中识别满足用户需求的信息，在正确的时间、地点和场景下仅向特定用户推送呢？

11月12日晚19:00，协会特邀中国移动天津公司AI实验室成员，也是CPDA数据分析师前辈刘静老师为大家带来《如何通过推荐系统实现业务增长》的主题沙龙活动。

近一个小时的沙龙活动中，刘老师给大家带来满满的干货，分享结束后刘老师和小伙伴们讨论热烈，频频互动。





## / 中国科学院全力建设可持续发展大数据国际研究中心 /

来源 / 中国网 编辑 / 协会会员处 李苗苗 日期 / 2020-09



中国网/中国发展门户网讯 9月22日，习近平主席在第75届联合国大会一般性辩论上发表重要讲话，宣布中国将设立可持续发展大数据国际研究中心，为落实《联合国2030年可持续发展议程》提供新助力。

今年是《联合国2030年可持续发展议程》通过5周年，在各国共同努力下，议程的全球落实取得重要进展，但也面临重重困难。突发的新冠肺炎疫情在全球蔓延，更加剧了各国实现可持续发展目标（SDGs）的挑战。

科技创新是实现可持续发展目标的重要手段。中国正在实施创新驱动战略，联合国提出技术促进机制，两者高度吻合，均是利用科学、技术和创新促进可持续发展。作为科技创新的重要方面，大数据在支撑可持续发展中具有重要作用。中国设立可持续发展大数据国际研究中心是服务《联合国2030年可持续发展议程》的一项重要举措。

可持续发展大数据国际研究中心是全球首个以大数据服务联合国可持续发展目标的机构。中心的目标是建成可持续发展大数据公共科技平台和研究机构，是可持续发展科学研究中心、数据信息服务与技术创新中心、人才培养与培训能力建设中心、全球可持续发展高端智库，服务联合国相关机构和会员国落实《联合国2030年可持续发展议程》。

据悉，可持续发展大数据国际研究中心将依托中国科学院建设。目前，中国科学院已集成优势研究力量和资源，系统深入地开展了科技支撑可持续发展的研究工作。2018年启动的中国科学院“地球大数据科学工程”专项，组织了来自

全国129个单位的1200余名专家，充分利用大数据、云计算、人工智能、空间技术、网络通信技术开展全球和区域多尺度可持续发展研究。中国科学院已实现了国家相关优势资源的整合，具备了中心启动和建设的条件。

中国科学院院长白春礼表示，可持续发展大数据国际研究中心的设立是我国助力落实《联合国2030年可持续发展议程》的具体行动。科学技术在推动实现可持续发展目标上的重要作用已成为国际共识。作为全球科技界的一员，中国科学院正在组织研究力量积极行动，加强科学-政策-社会的衔接互动，提供解决方案和科学支撑，助力推动《联合国2030年可持续发展议程》的全球落实，与国际科技界共同应对可持续发展面临的新挑战。



联合国技术促进机制10人组成员、“地球大数据科学工程”专项负责人郭华东院士指出，专项正在建设地球大数据云服务平台与数字地球科学平台，将发射服务可持续发展目标监测的地球科学卫星，发布了《地球大数据支撑可持续发展目标报告（2019）》，相关成果入列联合国可持续发展技术促进机制在线平台。可持续发展大数据国际研究中心的设立将为可持续发展目标提供有力的科技支撑。

# 国企数字化转型驶上快车道：组建国资大数据产业集团 设立大数据产业基金

来源 / 今晚报 编辑 / 协会会员处 李苗苗 日期 / 2020-11

天津市国有企业数字化转型论坛举行。据了解，为促进国企数字化转型，本市将组建国资大数据产业集团，设立大数据产业基金，同时融合公共服务行业数据，打造智慧津城。论坛期间，市国资委与华为公司签署了战略合作协议，双方将在工业互联网、大数据、云计算、人工智能、鲲鹏生态等领域开展全方位、深层次合作，共同推动本市国有企业数字化转型和产业升级，助力“五个现代化天津”建设。



党的十九届五中全会对加快数字化发展作出全面部署，提出“推进数字产业化和产业数字化，推动数字经济和实体经济深度融合，打造具有国际竞争力的数字产业集群”。市委、市政府布置实施的国企改革三年行动，明确提出要加速国企数字化、网络化、智能化转型升级。天津国有企业紧紧抓住产业数字化、数字产业化赋予的机遇，结合谋划“十四五”规划，加速推进新技术创新、新产品培育、新模式扩散和新业态发展，坚持数据驱动、集成创新、合作共赢的理念，尊重基层首创精神，持之以恒地做数字化转型的探索者、组织者、引领者、实践者。

津投资本党委书记、董事长周宏斌表示，津投资本坚持以产业发展为重心，打造“一个集团支撑、一只基金推动”，立足天津、面向京津冀、辐射全中国，力争成为全国领先的大数据产业生态建设典型案例。

一个集团支撑，即组建国资大数据产业集团。充分发挥

现有资源、能力优势，进一步完善IDC基础设施、系统集成、应用软件开发、业务运营服务、信息安全管理等基础能力建设，着力强化模型、算力、场景等大数据应用能力，充分发挥国资优势，积极布局智慧交通、智慧国资、智慧医疗、智慧公共服务等行业领域，打造协同创新、融合发展的产业新格局。

一只基金推动，即设立大数据产业基金。在大数据应用方面，遵循百花齐放原则，充分发挥市场化机制活力，激励大数据应用产业蓬勃发展，促进本市大数据产业迅速聚集，形成健康有序、充满活力的大数据产业生态。在此过程中，以天津本地的引导资金、国有资金为主，设立大数据产业基金，以市场化资本运作方式积极参与、扶持大数据应用项目在津落地发展。

市国资委对津投资本的另一个定位是公共服务类企业混改平台。公共服务行业为大数据的应用提供了丰富的业务场景，通过打通各个板块之间的“数据孤岛”，有效放大数据价值。对公交、地铁的用户大数据关联分析，可帮助优化线路设计，提升接驳便利度；通过全时的数据分析，可以改善运营管理。进一步将行业数据与交通管理部门、市政道路数据及静态交通数据融合，实时动态地对交通管理进行调整，优化交通管理水平，助力津城智慧交通。目前，津投资本已整合的中环系统、广通股份，在智慧城市领域均开展了富于特色的尝试。

据介绍，市国资委和本市国有企业于2019年启动了“智慧国资”系统的建设，目前“智慧国资”平台建设正在加快推进，已被纳入全国性国资国企在线监管系统，将有力提升服务国企转型发展的效能。

“天津国企将进一步加快数字化转型步伐，加快组织创新、技术创新、融合创新、跨界创新，积极发展数字产业、布局数字产业生态，在智能制造、智慧港口、智慧基础设施建设、智慧公共服务以及智慧物流、智慧金融、智慧旅游等方面培育更多新业务增长点。”市国资委党委书记、主任彭三表示。



# 《贵州省政府数据共享开放条例》12月1日起施行

来源 / 贵州人大 编辑 / 协会会员处 李苗苗 日期 / 2020-09

9月23日-25日，贵州省第十三届人民代表大会常务委员会第十九次会议在贵阳举行，会议表决通过了《贵州省政府数据共享开放条例》（以下简称《条例》），《条例》自2020年12月1日起施行。

《条例》共七章45条，从政府数据管理、政府数据共享、政府数据开放、监督管理等明确贵州省政府数据共享开放事项。官方表示，《条例》旨在推动政府数据共享开放，加快政府数据汇聚、融通、应用，培育发展数据要素市场，提升政府社会治理能力和公共服务水平，促进经济社会发展。

《条例》规定：“政府数据共享开放应当具有合法目的和用途，遵循正当、必要、适度的原则，依法维护国家安全、公共安全、经济安全、社会稳定，保守国家秘密，保护商业秘密、隐私和个人信息。”

《条例》明确，政府数据共享开放应当在全省统一的政府数据共享平台、开放平台上进行。省人民政府大数据主管部门负责全省统一的政府数据共享平台、开放平台的建设和运行维护。鼓励各市、州人民政府通过全省统一的政府数据共享平台建设本行政区域的基础数据、主题数据。

此外，《条例》还规定，任何单位或者个人对获取到的有条件共享开放的政府数据，不得擅自转让、改变用途或者使用范围，不得利用政府数据从事违法犯罪活动。

擅自转让获取的有条件开放的政府数据、改变用途或者使用范围的，由县级以上人民政府大数据主管部门责令改正，有违法所得的，没收违法所得，并可处以违法所得1倍以上10倍以下罚款；没有违法所得的，处以1万元以上5万元以下罚款。

## 贵州省政府数据共享开放条例

### 第一章 总则

第一条 为了推动政府数据共享开放，加快政府数据汇聚、融通、应用，培育发展数据要素市场，提升政府社会治理能力和公共服务水平，促进经济社会发展，根据有关法律、法规的规定，结合本省实际，制定本条例。

第二条 本省行政区域内政府数据共享开放及其相关管理

活动，应当遵守本条例。

涉及国家秘密的政府数据，按照相关保密法律、法规的规定执行。

第三条 本条例所称的政府数据，是指行政机关在依法履行职责过程中制作或者获取的，以一定形式记录、保存的各类数据，包括行政机关直接或者通过第三方依法采集、管理和因履行职责需要依托政府信息系统形成的数据。

本条例所称的政府数据共享，是指行政机关因履行职责需要使用其他行政机关政府数据和为其他行政机关提供政府数据的行为。

本条例所称的政府数据开放，是指行政机关面向公民、法人或者其他组织依法提供政府数据的行为。

第四条 政府数据共享开放应当具有合法目的和用途，遵循正当、必要、适度的原则，依法维护国家安全、公共安全、经济安全、社会稳定，保守国家秘密，保护商业秘密、个人信息和隐私。

第五条 省人民政府统一领导全省政府数据共享开放工作，建立议事协调机制，统筹协调政府数据共享开放工作的重大事项。市、州和县级人民政府负责本行政区域政府数据共享开放工作。

县级以上人民政府应当将政府数据共享开放工作纳入本行政区域国民经济和社会发展规划，相关工作所需经费纳入同级财政预算。

第六条 县级以上人民政府大数据主管部门负责本行政区域内政府数据共享开放的统筹管理和政府数据的协调调度等工作。

县级以上人民政府其他部门按照各自职责负责政府数据共享开放的相关工作。

第七条 使用各级财政资金建设和运行维护的政务信息系统，应当按照统一规划、统一建设、统一购买服务、统筹资金保障的方式进行建设和运行维护，促进数据汇聚、数据共享、系统互通、业务协同。

第八条 政府数据共享开放应当在全省统一的政府数据共享平台、开放平台上进行。

省人民政府大数据主管部门负责全省统一的政府数据共



享平台、开放平台的建设和运行维护。

## 第二章 政府数据管理

第九条 政府数据共享开放实行目录管理。省人民政府大数据主管部门应当制定全省政府数据共享目录和开放目录编制指南，定期发布政府数据共享、开放责任清单。

行政机关应当按照编制指南和责任清单要求，对照本单位职能编制、维护各自的全量政府数据共享目录和开放目录。编制目录时应当听取其他行政机关和社会公众的意见和建议，并报同级人民政府大数据主管部门审核。省人民政府大数据主管部门负责将各行政机关的共享目录和开放目录汇总后，在全省统一的政府数据共享平台、开放平台上发布。

行政机关应当对本单位的政府数据共享目录和开放目录定期进行更新，至少每年进行一次全面维护；因法律、法规调整或者职能变化需要更新目录的，应当自变化之日起15个工作日内完成。

第十条 人口信息、法人单位信息、自然资源和空间地理信息、电子证照等基础数据由省人民政府大数据主管部门会同相关行政机关归集、管理和维护，并通过政府数据共享平台在部门间进行无条件共享。

围绕经济社会发展的同一主题领域，由多部门共建项目形成的精准扶贫、卫生健康、社会救助、社会信用、生态环保、气象水文、食品安全、应急管理、城乡建设等主题数据，由主要负责的省级行政机关会同相关行政机关归集、管理和维护，并通过政府数据共享平台予以共享。

鼓励各市、州人民政府通过全省统一的政府数据共享平台归集、管理和维护本行政区域的基础数据、主题数据。

第十一条 行政机关应当依照法律、法规和部门职责、政府数据共享目录和开放目录及有关标准规范，及时采集、管理和维护政府数据，确保数据真实、准确、完整。

可以通过政府数据共享平台获得的政府数据，行政机关不得重复采集。

第十二条 行政机关应当对本部门共享开放数据设置标识，自然人信息应当以公民身份号码作为标识，法人及其他组织信息应当以统一社会信用代码作为标识。

第十三条 行政机关应当指定专门的数据管理机构或者专人负责本部门政府数据共享目录和开放目录编制以及数据的归集、发布、维护管理，负责配合政府数据共享开放协调调度工作。

## 第三章 政府数据共享

第十四条 政府数据以共享为原则、不共享为例外，按照共享属性分为无条件共享、有条件共享和不予共享三种类型。

可以提供给所有行政机关共享使用的政府数据属于无条件共享类。

可以提供给相关行政机关或者仅能够部分提供给其他行政机关共享使用的政府数据属于有条件共享类。

不宜提供给其他行政机关共享使用的政府数据属于不予共享类。

列入有条件共享类或者不予共享类的政府数据，应当有法律、行政法规或者国家有关规定作为依据。

第十五条 政府数据提供部门应当按照共享目录将共享数据及时、准确、完整的在政府数据共享平台上发布，并明确数据的共享范围和用途。

政府数据使用部门应当从政府数据共享平台上获取所需的数据，并根据履行职责的需要、数据提供部门明确的范围和用途使用共享数据。

第十六条 属于有条件共享类的政府数据，使用部门通过政府数据共享平台向数据提供部门提出申请，申请内容包括：

- (一) 申请部门的名称、联系方式；
- (二) 申请共享政府数据的理由；
- (三) 申请共享政府数据的类型和内容；

(四) 申请共享政府数据的用途、使用范围和安全管理措施等。

数据提供部门应当在10个工作日内予以答复，同意共享的，及时提供，使用部门按照答复意见使用共享数据；不同意共享的，数据提供部门应当说明理由。

第十七条 政府数据使用部门申请共享的数据未列入提供部门共享目录的，提供部门应当自收到申请之日起10个工作日内答复；同意共享的，应当列入目录进行采集，并按照规定实施共享；不同意共享的，应当说明理由。

第十八条 政府数据使用部门因履行职责确需使用不能共享的政府数据的，由使用部门与提供部门协商解决；不能达成一致意见的，由同级人民政府大数据主管部门协调解决。

第十九条 行政机关需要使用国家部委或者中央在黔单位政府数据的，可以提请省人民政府大数据主管部门与相关单位协调获取。

第二十条 行政机关通过政府数据共享平台提供的文书类、证照类等政府数据，加盖行政机关数据共享电子印章后，与纸质文书原件具有同等效力，可以作为其他行政机关履行行政管理职责的依据。

#### 第四章 政府数据开放

第二十一条 政府数据开放应当坚持需求导向、有序开放、平等利用、确保安全的原则，实行分类管理，按照开放属性分为无条件开放、有条件开放和不予开放三种类型。

可以提供给所有公民、法人或者其他组织使用的政府数据属于无条件开放类。

在特定条件下可以提供给公民、法人或者其他组织使用的政府数据属于有条件开放类。

涉及国家秘密、商业秘密、个人隐私，或者法律、法规规定不得开放的政府数据属于不予开放类。

列入有条件开放类或者不予开放类的政府数据，应当有法律、行政法规或者国家有关规定作为依据。

第二十二条 行政机关在开放政府数据前，应当对拟开放的敏感数据进行脱敏处理，防止泄露国家秘密、商业秘密、个人信息和隐私。

第二十三条 属于无条件开放类的政府数据，应当以可以机器读取的格式在政府数据开放平台发布，以便公民、法人或者其他组织获取、利用。

第二十四条 公民、法人或者其他组织需要使用有条件开

放的政府数据的，应当通过政府数据开放平台向数据提供部门提出申请，申请内容包括：

(一) 申请人的姓名或者名称、身份证明、联系方式；

(二) 申请开放政府数据的名称、类型、内容或者便于数据提供部门查询的其他特征性描述；

(三) 申请开放政府数据的用途、使用范围和安全管理措施等。

第二十五条 政府数据提供部门收到数据开放申请时，能够立即答复的，应当立即答复。

数据提供部门不能立即答复的，应当自收到申请之日起15个工作日内予以答复。如需要延长答复期限的，应当经数据提供部门负责人同意并告知申请人，延长的期限最长不得超过15个工作日。

数据提供部门同意政府数据开放申请的，通过政府数据开放平台及时向申请人开放，并明确数据的用途和使用范围；不同意开放的，应当说明理由。

第二十六条 申请人申请开放政府数据的数量、频次明显超过合理范围的，数据提供部门可以要求申请人说明理由。数据提供部门认为理由不合理的，告知申请人不予处理；数据提供部门认为理由合理的，应当及时向申请人开放。

第二十七条 县级以上人民政府及其大数据主管部门应当定期通过政府数据开放平台或者其他渠道加强政府数据开放的宣传和推广，收集公众对政府数据开放的意见建议，改进政府数据开放工作。

第二十八条 省人民政府应当建立政府数据资源有效流动和开发利用机制，推进政府数据资源的开发利用。

省和市、州人民政府应当公平择优选择具有相应管理经验、专业能力的法人或者其他组织，对非涉密但是涉及敏感信息的政府数据提供脱敏、清洗、加工、建模、分析等服务。

第二十九条 鼓励公民、法人或者其他组织利用政府数据资源创新产品、技术和服务，构建农业、工业、金融、交通、教育、城市管理、公共资源交易等领域规范化数据开发利用的场景，培育数字经济新产业、新业态和新模式，发挥政府数据资源的经济价值和社会效益。

第三十条 政府数据开发利用应当遵循合法、正当、安全的原则，不得损害国家利益、社会公共利益和第三方合法权益。

#### 第五章 监督管理

第三十一条 省人民政府大数据主管部门每年应当结合地

方发展特色、政务协同要求、企业利用需求和公众生活需要，明确政府数据共享开放的重点内容，指导和监督相关行政机关按照重点内容提供数据，拓展政府数据共享开放范围、深化共享开放内容、强化动态更新。

第三十二条 任何单位或者个人对获取到的有条件共享开放的政府数据，不得擅自转让、改变用途或者使用范围，不得利用政府数据从事违法犯罪活动。

第三十三条 县级以上人民政府及其部门应当加强政府数据共享开放的安全管理，做好政府数据保密审查和风险防范，定期开展安全培训、风险评估等工作。

行政机关为提升政府数据质量委托第三方开展政府数据规范治理等工作的，应当履行安全管理责任，通过签订保密协议、技术监测等措施，保障政府数据安全。

第三十四条 政府数据共享平台、开放平台应当具备数据调度溯源功能，调度与操作记录应当长期保存，每年定期进行安全检测评估。

第三十五条 法人或者其他组织对政府数据开展脱敏、清洗、加工、建模、分析等数据服务时，应当构建安全可靠的数据开发环境，依照法律、法规履行安全管理责任，保护政府数据免受泄露、窃取、篡改、毁损与非法使用。

第三十六条 省和市、州人民政府大数据主管部门负责对本行政区域政府数据的开发利用实施统一监督管理。

提供数据服务的法人或者其他组织通过对省和市、州人民政府的数据开发利用得出的数据模型、数据产品等，应当报同级人民政府大数据主管部门进行安全审查。

提供数据服务的法人或者其他组织需要与第三方合作进行政府数据开发利用的，应当按照安全管理要求签订合作开发协议，并报同级人民政府大数据主管部门备案。

第三十七条 建立政府数据使用反馈机制。

使用政府数据的单位或者个人对获取的政府数据发现不完整或者有错误的，可以向数据提供部门反馈，数据提供部门应当及时补充、校核和更正。

政府数据共享时，数据提供部门有权了解数据使用部门使用相关数据的情况。

第三十八条 省人民政府大数据主管部门制定政府数据共享开放工作考核评价标准。县级以上人民政府应当根据考核评价标准，每年对本级行政机关、下级人民政府数据共享目录和开放目录的维护管理、数据采集与更新、数据共享开放、数据

使用、数据开发利用效果等情况进行考核评价，定期通报评价结果并纳入年度目标考核；还可以委托第三方对政府数据共享开放的程度和效果进行评估，结果向社会公布。

## 第六章 法律责任

第三十九条 行政机关及其工作人员违反本条例规定，有下列行为之一，由上级机关，主管部门，任免机关、单位或者监察机关责令改正；情节严重的，对直接负责的主管人员和其他直接责任人员依法予以处分：

（一）未按照要求编制或者发布本部门政府数据共享目录或者开放目录、不及时发布或者更新政府数据的；

（二）重复采集可以从政府数据共享平台获取的政府数据的；

（三）拒不答复政府数据共享、开放申请或者未按照要求提供数据的；

（四）对已经发现不完整或者有错误的数据库，拒不进行补充、校核和更正的；

（五）未落实政府数据安全职责的；

（六）擅自改变政府数据用途和使用范围的；

（七）违反本条例规定的其他行为。

第四十条 行政机关按照法律、法规的规定开展政府数据共享、开放，并履行了监督管理职责和合理注意义务的，依法不承担或者免于承担因共享开放的数据质量等问题产生的相应责任。

第四十一条 公民、法人或者其他组织违反本条例规定，擅自转让获取的有条件开放的政府数据、改变用途或者使用范围的，由县级以上人民政府大数据主管部门责令改正，有违法所得的，没收违法所得，并可处以违法所得1倍以上10倍以下罚款；没有违法所得的，处以1万元以上5万元以下罚款。

第四十二条 违反本条例规定的其他行为，法律、法规有处罚规定的，从其规定。

## 第七章 附则

第四十三条 法律、法规授权具有公共事务管理职能的组织实施的数据共享开放活动，适用本条例。

第四十四条 教育、卫生健康、供水、供电、供气、供热、环境保护、公共交通等与人民群众利益密切相关的公共企事业单位开展数据共享开放活动，依照相关法律、法规和国务院、省有关主管部门或者机构的规定执行。

第四十五条 本条例自2020年12月1日起施行。

# 深圳综合改革试点实施方案发布 支持建设粤港澳大湾区数据平台

来源 / CCTV 编辑 / 协会会员处 李苗苗 日期 / 2020-09

近日，中共中央办公厅、国务院办公厅印发了《深圳建设中国特色社会主义先行示范区综合改革试点实施方案（2020—2025年）》，并发出通知，要求各地区各部门结合实际认真贯彻落实。

其中《方案》提及，加快培育数据要素市场。要求率先完善数据产权制度，探索数据产权保护和利用新机制，建立数据隐私保护制度。试点推进政府数据开放共享。支持建设粤港澳大湾区数据平台，研究论证设立数据交易市场或依托现有交易场所开展数据交易。开展数据生产要素统计核算试点。

《深圳建设中国特色社会主义先行示范区综合改革试点实施方案（2020—2025年）》主要内容如下：

党中央作出兴办经济特区重大战略部署40年来，深圳敢闯敢试、敢为人先、埋头苦干，创造了发展史上的奇迹，成为全国改革开放的一面旗帜。

2019年8月，以习近平同志为核心的党中央作出支持深圳建设中国特色社会主义先行示范区的重大决策，一年来，各项工作取得积极进展。以设立经济特区40周年为契机，在中央改革顶层设计和战略部署下，支持深圳实施综合授权改革试点，是新时代推动深圳改革开放再出发的又一重大举措，是建设中国特色社会主义先行示范区的关键一招，也是改革创新方式方法的全新探索，为贯彻落实习近平总书记关于深圳改革发展的重要指示批示精神和《中共中央、国务院关于支持深圳建设中国特色社会主义先行示范区的意见》有关要求，积极稳妥做好综合授权改革试点工作，制定本方案。

## 一、总体要求

（一）指导思想。以习近平新时代中国特色社会主义思想为指导，全面贯彻党的十九大和十九届二中、三中全会精神，按照党中央、国务院决策部署，牢牢把握正确改革方向，围绕中国特色社会主义先行示范区的战略定位和战略目标，赋予深圳在重点领域和关键环节改革上更多自主权，支持深圳在更高起点、更高层次、更高目标上推进改革开放，率先完善各方面制度，构建高质量发展体制机制，推进治理体系和治理能力现代化，加快形成全面深化改革、全面扩大开放新格局，推动更高水平深港合作，增强在粤港澳大湾区建设中的核

心引擎功能，努力创建社会主义现代化强国的城市范例。

## （二）工作原则

——坚持解放思想、守正创新。坚持和加强党的全面领导，坚持中国特色社会主义道路，坚持以人民为中心的发展思想，在守正的基础上创新，解放思想、实事求是，固根基、扬优势、补短板、强弱项。

——坚持市场化、法治化、国际化。充分发挥市场在资源配置中的决定性作用，更好发挥政府作用，推进改革与法治双轮驱动，实施更大范围、更广领域、更深层次的全面开放。

——坚持系统集成、协同高效。突出改革系统性、整体性、协同性，聚焦重点领域和关键环节，推动各方面制度更加衔接配套、成熟定型，实现改革目标集成、政策集成、效果集成。

——坚持先行先试、引领示范。坚持问题导向、目标导向、结果导向，给予充分改革探索空间，在遵循宪法和法律、行政法规基本原则前提下，允许深圳立足改革创新实践需要，根据授权开展相关试点试验示范。鼓励大胆创新、真抓实干，注重经验总结，及时规范提升，为全国提供示范。

——坚持底线思维、稳步实施。提前预设底线情形，在风险总体可控前提下，科学把握时序、节奏和步骤，分期分步、稳妥有序推进改革。

（三）主要目标。2020年，在要素市场化配置、营商环境优化、城市空间统筹利用等重要领域推出一批重大改革措施，制定实施首批综合授权事项清单，推动试点开好局、起好步。2022年，各方面制度建设取得重要进展，形成一批可复制可推广的重大制度成果，试点取得阶段性成效。2025年，重要领域和关键环节改革取得标志性成果，基本完成试点改革任务，为全国制度建设作出重要示范。

## 二、完善要素市场化配置体制机制

（四）支持在土地管理制度上深化探索。将国务院可以授权的永久基本农田以外的农用地转为建设用地审批事项委托深圳市政府批准。支持在符合国土空间规划要求的前提下，推进二三产业混合用地。支持盘活利用存量工业用地，探索解决规划调整、土地供应、收益分配、历史遗留用地问题。探索利用存量建设用地进行开发建设的市场化机制，完善闲置土地使用权收回机制。深化深汕特别合作区等区域农村土地制度改

革。支持依托公共资源交易平台建设自然资源资产交易市场，完善一二级市场联动的土地市场服务监管体系。试点实行土地二级市场预告登记转让制度。

(五) 完善适应超大城市特点的劳动力流动制度。深化户籍制度改革，调整完善积分落户政策。完善居住证制度，鼓励根据实际扩大公共服务范围、提高服务标准，稳步推进基本公共服务常住人口全覆盖。允许探索适应新技术、新业态、新产业、新模式发展需要的特殊工时管理制度。

(六) 支持在资本市场建设上先行先试。推进创业板改革并试点注册制，试点创新企业境内发行股票或存托凭证(CDR)。建立新三板挂牌公司转板上市机制。优化私募基金市场准入环境。探索优化创业投资企业市场准入和发展环境。依法依规开展基础设施领域不动产投资信托基金试点。在中国人民银行数字货币研究所深圳下属机构的基础上成立金融科技创新平台。支持开展数字人民币内部封闭试点测试，推动数字人民币的研发应用和国际合作。

(七) 加快完善技术成果转化相关制度。改革科研项目立项和组织方式，建立主要由市场决定的科技项目遴选、经费分配、成果评价机制。深化科技成果使用权、处置权和收益权改革，在探索赋予科研人员职务科技成果所有权或长期使用权、成果评价、收益分配等方面先行先试。探索政府资助项目科技成果专利权向发明人或设计人、中小企业转让和利益分配机制，健全国有企业科技成果转化利益分配机制。完善技术成果转化公开交易与监管体系。

(八) 加快培育数据要素市场。率先完善数据产权制度，探索数据产权保护和利用新机制，建立数据隐私保护制度。试点推进政府数据开放共享。支持建设粤港澳大湾区数据平台，研究论证设立数据交易市场或依托现有交易场所开展数据交易。开展数据生产要素统计核算试点。

(九) 健全要素市场评价贡献机制。率先探索完善生产要素由市场评价贡献、按贡献决定报酬的机制，增加劳动者特别是一线劳动者劳动报酬。充分尊重科研、技术、管理人才，探索充分体现技术、知识、管理、数据等要素价值的实现形式。深入推进区域性国资国企综合改革试验，支持建立和完善符合市场经济规律与企业家成长规律的国有企业领导人员管理机制，探索与企业市场地位和业绩贡献相匹配、与考核结果紧密挂钩、增量业绩决定增量激励的薪酬分配和长效激励约束机制。

### 三、打造市场化法治化国际化营商环境

(十) 进一步完善公平开放的市场环境。在全国统一的市场准入负面清单基础上，制定深圳放宽市场准入特别措施清单，放宽能源、电信、公用事业、交通运输、教育等领域市场

准入。进一步放宽前沿技术领域的外商投资准入限制。支持完善公平竞争制度。完善经营邮政通信业务审批机制。试点能源领域许可审批告知承诺制。推进破产制度和机制的综合配套改革，试行破产预重整制度，完善自然人破产制度。

(十一) 打造保护知识产权标杆城市。开展新型知识产权保护试点，完善互联网信息等数字知识产权财产权益保护制度，探索建立健全证据披露、证据妨碍排除和优势证据规则，建立知识产权侵权惩罚性赔偿制度。探索在部分知识产权案件中实行举证责任转移制度。实施知识产权领域以信用为基础的分级分类监管。

(十二) 完善行政管理体制和经济特区立法。按程序赋予深圳在干部和机构管理、统筹使用各类编制资源等方面更大自主权。探索完善行政争议多元解决机制，健全行政复议与行政诉讼衔接机制。支持深圳扩宽经济特区立法空间，在新兴领域加强立法探索，依法制定经济特区法规规章。

## 四、完善科技创新环境制度

(十三) 优化创新资源配置方式和管理机制。支持实行非竞争性、竞争性“双轨制”科研经费投入机制。推动完善科研机构管理机制，建立常态化的政企科技创新咨询制度。实施高层次科技人才定向培养机制。

(十四) 建立具有国际竞争力的引才用才制度。按程序赋予深圳外国高端人才确认函权限，探索优化外国人来华工作许可和工作类居留许可审批流程。支持探索制定外籍“高精尖缺”人才认定标准，为符合条件的外籍人员办理R字签证和提供出入境便利。为符合条件的外籍高层次人才申请永久居留提供便利。支持探索建立高度便利化的境外专业人才执业制度，放宽境外人员(不包括医疗卫生人员)参加各类职业资格考试的限制。

## 五、完善高水平开放型经济体制

(十五) 加大制度型开放力度。支持以规则衔接深化粤港澳大湾区合作发展。充分发挥中国(广东)自由贸易试验区深圳前海蛇口片区全面深化改革和扩大开放试验田作用，形成更多可复制可推广的制度创新成果。加强对重大疑难涉外商事案件的业务指导。支持完善法治领域跨境协作机制，健全国际法律服务和纠纷解决机制。

(十六) 扩大金融业、航运业等对外开放。支持符合条件的在深境内企业赴境外上市融资。开展本外币合一跨境资金池业务试点。支持深圳在推进人民币国际化方面先行先试，推动完善外汇管理体制。支持符合条件的外资金融机构在深圳依法发起设立证券公司、基金管理公司。支持符合条件的外资机构在深圳依法合规获取支付业务许可证。推动构建与国际接轨

的金融规则体系。探索完善国际船舶登记制度。赋予深圳国际航行船舶保税加油许可权，进一步放开保税燃料油供应市场。

## 六、完善民生服务供给体制

(十七) 创新医疗服务体系。支持在深圳开展国际前沿药品临床应用。探索完善医疗服务跨境衔接，建立与国际接轨的医学人才培养、医院评审认证标准体系。支持建设全新机制的医学科学院。支持完善重大疫情防控体制机制，健全公共卫生应急管理体系。

(十八) 探索扩大办学自主权。探索扩大在深高等学校办学自主权。在符合国家相关政策规定前提下，支持深圳引进境外优质教育资源，开展高水平中外合作办学。赋予深圳对企业博士后科研工作站分站的设立和撤销权限。

(十九) 优化社会保障机制。探索公共服务多元化供给新机制，加快构建以促进健康为导向的创新型医保制度。支持完善普惠婴幼儿照护服务体系。鼓励利用全国一体化政务服务平台实现医保政务服务一体化办理。支持健全公共就业服务和终身职业技能培训制度。

(二十) 完善文化体育运营管理体制。支持深化文艺院团改革，完善院团管理体制、运行机制和利益分配制度。按程序赋予省级电视剧审查等管理权限。支持建设适用国际通用规则的文化艺术品（非文物）拍卖中心。支持开展体育消费城市试点，推进体育产业创新试验，创新促进体育赛事发展的服务管理机制和安保制度。

## 七、完善生态环境和城市空间治理体制

(二十一) 健全生态建设和环境保护制度。支持完善生态保护红线、环境质量底线、资源利用上线和生态环境准入清单等“三线一单”生态环境分区管控体系，开展重要生态空间自然资源确权登记，扩大生态系统服务价值核算范围。支持完善本地清洁能源供应机制，建设能源产业创新中心、创新联合体等平台机构。支持完善产品环保强制性地方标准，建立绿色产业认定规则体系，完善气候投融资机制。推动完善陆海统筹的海洋生态环境保护修复机制，实行环境污染强制责任保险制度，探索建立入海排污口分类管理制度。加快建设国家可持续发展议程创新示范区。

(二十二) 提升城市空间统筹管理水平。按照相关法律法规，支持推动在建设用地上、地表和地下分别设立使用权，探索按照海域的水面、水体、海床、底土分别设立使用权，促进空间合理开发利用。开展深化自然生态空间用途管制试点。按程序赋予深圳占用林地省级审核权限。探索优化用地用林用海“统一收文、统一办理、统一发文”审批机制，推动自然资源使用审批全链条融合。开展航空资源结构化改革试

点。完善无人机飞行管理制度。

## 八、强化保障措施

(二十三) 全面加强党的领导。坚持和加强党对深圳建设中国特色社会主义先行示范区综合改革试点的领导，增强“四个意识”、坚定“四个自信”、做到“两个维护”，把党的建设始终贯穿综合改革试点全过程，严密党的组织体系，为深圳深化改革开放提供坚强保障。

(二十四) 创新工作机制。支持深圳结合实际率先开展相关试点试验示范，实施重大改革举措。国家发展改革委会同有关方面分批次研究制定授权事项清单，按照批量授权方式，按程序报批后推进实施。有关方面要按照本方案要求和经批准的事项清单，依法依规赋予深圳相关管理权限。建立健全重大风险识别及系统性风险防范制度和政策体系。实行包容审慎的改革风险分类分级管控机制，根据风险程度，分别采取调整、暂缓或终止等处置措施，不断优化综合改革试点实施路径。国家发展改革委会同有关部门及时跟进综合改革试点进展，加强统筹协调和指导评估，对达到预期效果的抓紧总结推广，对新情况新问题及时分析评估，重要情况及时向党中央、国务院报告。

(二十五) 落实地方责任。广东省要积极为深圳开展综合改革试点创造条件，加大行政审批、科技创新、规划管理、综合监管、涉外机构和组织管理等方面放权力度，依法依规赋予深圳更多省级经济社会管理权限。深圳要切实担负起试点主体责任，增强使命感和责任感，认真做好具体实施工作，确保各项改革任务扎实有序推进。要加强与粤港澳大湾区其他城市协调合作，建立区域互动、优势互补的改革联动机制，实现协同对接，充分发挥制度整体效能。

(二十六) 强化法治保障。建立健全与综合改革试点相配套的法律法规、政策调整机制。本方案提出的各项改革政策举措，凡涉及调整现行法律或行政法规的，经全国人大常委会或国务院授权后实施。有关方面要加强指导和服务，统筹综合改革试点涉及的法律法规事项，做好与相关法律法规立改废释的衔接。

(二十七) 营造改革氛围。弘扬特区精神，继续大胆地闯、大胆地试。健全改革的正向激励机制，注重在改革一线考察识别干部，大胆提拔使用敢于改革、善于改革的干部，充分调动和激发广大干部参与改革的积极性、主动性和创造性。全面落实“三个区分开来”，建立健全容错纠错机制，宽容干部在改革创新中的失误错误，对干部的失误错误进行综合分析，该容的大胆容，不该容的坚决不容。及时宣传深圳推进改革的新进展新成效，为综合改革试点营造良好舆论氛围。

## / 让"大数据+医疗"更好满足群众需求 /

来源 / 贵阳日报 编辑 / 协会会员处 李苗苗 日期 / 2020-09



电子病历、手机预约挂号、资深专家线上会诊……近年来，贵阳市深入推进大数据与医疗深度融合，让优质医疗资源触“屏”可及。特别是在新冠肺炎疫情防控的背景下，互联网医疗给群众带来了许多便利，使我们深感“大数据+医疗”的诸多新探索，让个性化、精准化的医疗服务有了更多可能。

近年来，医疗卫生行业IT市场年增长率保持在14%以上，移动医疗市场增速高达20%以上，已具有丰富的健康医疗大数据应用发展实践。“大数据+医疗”让健康数据“多跑路”、人民群众“少跑腿”，健康咨询、预约就诊、诊间结算等给老百姓带来了更加便捷的应用服务；让优质医疗资源“下得去”，更好地推动分级诊疗落地，加快远程医疗普及，推动精准医疗发展；让健康管理更有招，更有利于做好预防为主、防治结合，通过可穿戴设备等做好慢病管理，及时监控血压、心率等方面的生命体征指标，同时进行健康提醒。

“大数据+医疗”既是一项重大民生工程，可以满足群众需求，也能培育新业态、形成新的经济增长点。在数据采集利用方面，可为现有医院向“互联网医院”转型提供帮助，通过场景应用等获取一部分数据，可与更多优质可穿戴设备厂商合作，实现互惠互利；

在数据分析、服务输出方面，可进一步加强与卫生主管部门合作，共同进行医疗大数据的分析和应用研究，可不断完善建设多维度一体化的服务链条，包括网上疾病自诊自查、在线问诊、预约挂号、线下陪诊陪护、院后随诊、医药电商、可穿戴设备云健康管理等服务，将大健康产业做得越来越大。

如今，老百姓的健康意识越来越强，对医疗健康的需求也越来越高。在加快推进“中国数谷”建设的征程中，我们应瞄准人民群众的现实需求，有效化解社会服务难点痛点问题，以“大数据+”拉近社会服务与公众之间的距离，在增进民生福祉中提升人民群众的幸福指数，让数字红利惠及广大人民群众。



## / 大数据风控将助推网贷中介平台转型 /

来源 / 中国经济新闻网 编辑 / 协会会员处 李苗苗 日期 / 2020-09

近年来，人工智能、区块链、云计算、大数据等新技术的不断涌现，越来越多从事助贷行业的金融科技公司意识到，大数据作为当前金融科技的重要依托，在风控体系中发挥着至关重要的作用，而大数据风控也在近几年成为了各大金融科技企业争相布局的重要据点。毋庸置疑的是，大数据风控不仅能有效提高金融服务的效率和安全性，降低风控成本，还能促进风险管理差异化和业务人性化，在金融科技业中有着重要作用。

尤其是今年7月17日发布了《商业银行互联网贷款管理暂行办法》(以下简称《暂行办法》)。《暂行办法》的出台，让助贷业务获得监管认可，进入合规发展通道，同时政策也允许围绕线上贷款业务的其他机构，如大数据、风控建模、智能风控等公司，在暂行办法条款约束之下健康地发展。可以说，2020年是助贷业务合规发展的开局之年，行业即将进入“黄金发展期”。

首先，从互联网金融开始，线上借贷业务经历了近十年发展已经趋于成熟化。并且互联网金融催生出了“大数据风控”等技术。再者，对于那些自身具有一定技术能力的互联网金融平台来讲，合规发展助贷业务、金融科技业务，在当下整个宏观政策指导的背景下，正是转型的最佳时机。

其次，大数据风控是当前网络借贷中介结构转型升级的重要一环。当前的实时大数据处理平台可以为其产品运营、风控、营销预测等场景提供全数字化支撑。利用领先的人工智能技术，可在秒级内对数据进行整合分析，从而计算出客户的风险程度。根据相关政策规定，遵照“法治化、市场化、信息中介化”的原则，从事网络借贷信息中介相关业务，以互联网为渠道，“为出借人与借款人提供直接借贷信息的采集整理、甄别筛选、网上发布，以及资信评估、借贷撮合、融资咨询、在线争议解决等相关服务”，以实现出借人与借款人的“点对点”借贷服务。

平台业务以“小额分散”的业务规模为主，出借人自行选择借款标的，与借款人实现点对点交易，建立民间借贷法律关系。有了大数据的介入，中间各个环节就一目了然，责权利明晰。

值得关注的是，为了打击“恶意逃废债”现象，平台全



再次，随着大数据风控的发展，金融科技将服务于更广泛的消费客群，技术对金融的赋能效力全社会有目共睹。考虑到部分出借人在借款标的尚未到期时有资金使用需求，平台开通债权转让功能，与出借人在合同中明确约定，只有在服务期满后，出借人才可发起债转申请。出借人申请债转时，其持有的借款标的并未到期，出借人申请债转的实质是为提前债权转让并退出平台服务。

平台作为网络借贷信息中介服务机构在出借人申请债转过程中的角色只是信息中介，在出借人申请债权转让时最大化协助出借人寻找债权受让方，如果有其他出借人愿意受让该债权，则转让成功。如果转让不成功，则出借人需继续持有借款标的的到期。对于出借人的债转申请，平台在官网、APP及出借人签署的相关协议中均明确表示不承诺保障债转成功。另外，债权转让是否成功与兑付无关。

随着国家对金融科技的愈发重视，监管的加强，正在让金融科技以迅雷不及掩耳之势飞速的发展与完善，部分金融科技平台也将牢牢把握金融科技数字化转型的关键趋向，完善大数据风控，为用户打造金融新体验。

## / 善用大数据让老赖没法赖 /

来源 / 广州日报 编辑 / 协会会员处 李苗苗 日期 / 2020-09



日前，广州互联网法院使出对付老赖的撒手锏——“E链云镜”智能执行分析系统，依托互联网大数据、人工智能技术，深挖被执行人的网络活动轨迹及财产线索，对其综合履行能力来个大数据评估“画像”，让那些有能力却拒不履行的老赖无所遁形。

“E链云镜”系统的上线，是运用“黑科技”破解执行难的一个重要突破口。现实中，法院在遭遇老赖时常常面临执行难问题，因为老赖为了逃避执行，往往千方百计隐藏和转移财产，营造出“缺乏履行能力”“无财产可供执行”的假象。随着互联网时代的到来，人们的生活方式、财产形式及转移方式等也全面“互联网化”，法院仅靠银行账户余额等财产静态余额的查控分析，已经不足以区分是“拒不履行”还是真的“执行不能”。而“E链云镜”系统通过与多家互联网企业建立司法协助执行机制，能及时调取被执行人一定时期的网上理财、网购消费、直播打赏等数据，从而对被执行人精准“画像”，清晰呈现其债务偿还履行能力的整体评估情况。

“E链云镜”的核心在于善用大数据。而要让大数据发挥更大的威力，关键还在一个“联”字。其一，要不断扩展

“联”的广度。如接入更多互联网平台的数据库，通过多维度数据更加精准地识别老赖，以此提高司法效率和办案质量。其二，要不断挖掘“联”的深度。这一点在实施联合惩戒方面尤为重要。解决执行难不能仅靠法院单打独斗，更需要争取包括其他部门、行业组织等在内的尽可能多的支持，通过联合全社会力量，共同惩戒失信行为，倒逼老赖主动履行义务。

目前，联合惩戒已颇见成效。老赖一旦被纳入失信名单，大到开公司、做高管，小到坐飞机、高铁时的位置，都会处处受限。而要织密惩戒失信的天罗地网，最大限度地压缩老赖生存空间，还需要在“联”的深度上下更多功夫。如“E链云镜”的动态预警功能，就是比较好的探索，其通过分析被执行人的经济活动数据，可以对非正常大额交易行为和疑似转移财产行为进行提示预警，这对于相关单位及时配合法院冻结有关资金极为重要。

只有“联”好大数据，从法律、经济、生活等各方面把老赖“困”住，让其无立锥之地，才能促使诚实守信、敬畏法律成为每个人的自觉。

## / 大数据技术为制造业提供多方位、精细化的服务 /

来源 / 智能制造网 编辑 / 协会会员处 李苗苗 日期 / 2020-09



大数据在制造业行业的应用，囊括了诸多细小的场景。来源于产品生命周期的各个环节，包括设计、制造、服务、市场、再利用各个环节，每个环节都会有大数据。“全”生命周期汇合起来的数据既多元又复杂。有了大数据技术，制造业得以加“数”前进。

### 供给链

现代供应链正在演变，并变得越来越复杂。大数据分析解决方案可提供供应链可视性，即时了解关键供应链的各种信息，例如哪些供应商表现良好，他们的产品制造种类、周期、质量等情况，以及按时交付订单的能力有多强。

在大数据技术的支持下，制造业企业可以为每一件产品标记一个专有的射频识别码，并通过这一射频识别码记录该产品在整个生产流程中的生产、包装等数据。一旦出现次品，工程师们可以通过这些数据排查整个生产环节，迅速找出生产线的缺陷并加以改进，进而大幅降低次品率。

RFID等产品电子标识技术、物联网感知技术以及终端云共享技术，能帮助制造业企业获得完整的产品供应链大数

据。工厂管理者利用这些数据进行分析，可以积极改变产品的制造、流通模式，以促进仓储、配送、销售效率的提升并降低成本。



### 需求链

大数据是一个很好的销售分析工具。通过历史数据的多维度组合，可以看出区域性需求占比和变化、产品品类的市场

受欢迎程度以及产品营销认可度较高的组合形式、消费者的需求变动等，以此来调整产品策略和铺货策略。

在需求链环节，大数据分析在为企业带来更多价值的同时，也对企业的技能型人才提出了更高的要求。由于大数据分析非传统的数据分析，传统数据分析可能依托于表格进行分析，而大数据分析是对杂乱的、海量的、杂乱的、多元的数据进行专业的分析，这就需要有专业的人员和去进行数据的整合、分析和调用，从而充分发挥各类数据的价值。

在大数据技术支持下，企业还可以利用传感技术、自动化技术等增强产品生产的智能性、网络性，将传统制造业和高端服务业融合在一起，进一步提高企业产品的竞争力。大数据能够为制造业提供多方位、精细化的服务，从产品设计到制造、从使用到维护、从在线推广到线下展示阶段，多元化正向数据以及逆向数据，都将在制造业供应链和需求链场景下得到全面应用，智慧工厂、智能机器人、智慧仓储等应用也为时不远。

### 总结：

大数据技术能够对海量数据信息进行搜集、统计、分析和处理，为人们的信息反馈、城市建设、商业活动、公共决策等提供重要参考，可以被广泛应用于金融、商业、教育、医疗、管理、电子等各个领域。

不过，面对数据孤岛严重、数据采集方式落后、缺乏统一数据资源管理机制、大数据行业应用标准缺失等问题，应该引起各方面的重视，尤其应加紧对数据隐私、数据安全方面的立法和规范。

以数据挖掘、数据分析为核心的应用和服务，无疑能为社会经济增长奠定坚实的数据基础。展望未来，大数据不仅会在不同层面改变大家的思维模式，还能改变许多人原有的生活方式和商业服务业态。

## / 治理“大数据杀熟”重在对症下药 /

来源 / 央广网 编辑 / 协会会员处 李苗苗 日期 / 2020-09

治理“大数据杀熟”现象，不是要“杀死”大数据，而是要善于借力，形成监管和治理的合力。监管部门应建立和完善大数据网上监管平台，提高对各种隐性“大数据利用”违法行为的查处能力。要将消费评价权保障、旅游者信息使用等纳入重点监管和治理范畴，与时俱进升级监管手段，打造让消费者“说走就走”、安全旅游的法治环境。

国庆节将至，一条与在线旅游相关的话题——“大数据杀熟行为10月1日起明令禁止”登上微博热搜榜。该话题缘于文化和旅游部印发的《在线旅游经营服务管理暂行规定》（简称《规定》）今年10月1日起正式施行，《规定》第十五条明确，在线旅游经营者不得滥用大数据分析等技术手段，基于旅游者消费记录、旅游偏好等设置不公平的交易条件，侵犯旅游者合法权益。这条规定针对的，就是近年来饱受诟病的“大数据杀熟”行为。

“大数据时代”到来，给人们带来了诸多便利，同时也

带来了一些负面影响，“大数据杀熟”便是其中之一。去年10月文化和旅游部发布的《在线旅游经营服务管理暂行规定》（征求意见稿）中，禁止“大数据杀熟”被定义为“在线旅游经营者不得利用大数据等技术手段，针对不同消费特征旅游者，对同一产品或服务在相同条件下设置差异化的价格”。

综合《规定》（征求意见稿）和《规定》的这两条规定，“大数据杀熟”可以简单理解为：在线旅游经营者滥用大数据分析手段，利用自身掌握的信息优势，对老客户设置比新客户更高的价格，造成老客户吃亏。去年3月，北京市消协发布的一项调查结果显示，近九成被调查者认为“大数据杀熟”现象普遍存在，56.92%的被调查者表示有过被“大数据杀熟”的经历，其中网购、在线旅游、酒店住宿、网约车、外卖、影视等消费场景最容易被“大数据杀熟”。

俗话说“人熟为宝”，在传统的商业道德里，熟客是应



该享受优待的。遗憾的是，在大数据时代，一些商家却反其道而行之，搞起了“大数据杀熟”，专门让熟客吃亏。所谓“最懂你的人伤你最深”，在“大数据杀熟”的问题上体现得淋漓尽致。当“德治”行之有限时，“法治”就应当站出来兜底，《在线旅游经营服务管理暂行规定》的出炉，便契合了这样的现实语境。《规定》明确禁止“大数据杀熟”，就是法律法规与时俱进保障消费者权益的写照，也是法治进步的佐证。

禁止“大数据杀熟”，徒法不足以自行，关键还得看落实。就眼下而言，一些现实问题亟待破解。一方面，处罚力度似显不足。《规定》第二十四条提出，文旅部门对有不诚信经营、滥用技术手段设置不公平交易条件等违法违规经营行为的在线旅游经营者，可通过约谈等行政指导方式予以提醒、警示、制止，并责令其限期整改。有的“大数据杀熟”实施时间较长且涉及众多消费者，如果仅通过行政指导方式进行处理，其处罚和震慑力度显然是不够的。

另一方面，“大数据杀熟”的判定标准虽已明确，但如何举证、谁来举证仍是大问题。消费者与经营者就“大数据杀熟”发生纠纷，有必要实行“举证倒置”，即由在线旅游平台

等经营者自证清白。对在线旅游经营者的各种说辞，如“产品本来就具有弹性空间”“存在支付方式、是否含早、取消政策、不同供应商等原因”“新客户是优惠券折扣后的价格”等，监管部门也不能听之任之，而必须依法严格核实查处。

“大数据杀熟”之所以存在，不仅在于平台拥有数据优势，也在于市场开放性和透明度不足。治理“大数据杀熟”现象，不是要“杀死”大数据，而是要善于借力，形成监管和治理的合力。监管部门应建立和完善大数据网上监管平台，提高对各种隐性“大数据利用”违法行为的查处能力。要积极落实《规定》的要求，将消费评价权保障、旅游者信息使用等纳入重点监管和治理范畴，与时俱进升级监管手段，切实维护消费者合法权益，打造让消费者“说走就走”、安全旅游的法治环境。

# / 深度抽吸对卷烟烟气指标影响的研究 /

作者 / 王艳丽, 史中华, 金玉善, 杨福利 编辑 / 协会会员处 李苗苗 日期 / 2020-09

**摘要:** 为了确定采用新的卷烟抽吸实验的ISO方法(深度抽吸方法)后,我们卷烟厂系列卷烟的常规烟气指标(焦油、烟碱、CO量)实测值的变化情况,我们将采用加拿大健康署的深度抽吸方法和国际标准方法对我们卷烟厂系列卷烟的常规烟气指标进行测定和比较,讨论抽吸方法的改变对我们卷烟厂系列卷烟烟气释放量的影响。

**关键词:** 卷烟; 加拿大深度抽吸模式; 烟气化学成分; 影响; 思考。

人类吸烟行为千变万化,没有任何一种抽吸模式能完全模拟人类吸烟行为。目前大部分国家和地区根据国际标准ISO 4387 测试烟气释放量并制定限量法规,比如欧盟规定自2004年1月1日起烟气焦油量不高于10 mg/支,烟气烟碱量不高于1 mg/支,烟气一氧化碳(CO)量不高于10 mg/支。随着全球控烟运动的不断高涨,特别是《烟草控制框架公约》(FCTC)的正式生效(2005年),以世界卫生组织(WHO)为代表的公共卫生组织对ISO/FTC抽吸模式提出质疑,认为ISO/FTC抽吸模式测试得到的烟气释放量低于人类实际烟气摄入量,并推荐制定“加拿大深度抽吸模式”国际标准。卷烟抽吸模式已成为世界烟草行业和公共卫生部门研究和争论的焦点之一。

世界卫生组织(WHO)明确提出要改变现行的卷烟抽吸实验的国际标准方法(ISO方法),取而代之以最严苛的深度抽吸方法:加拿大健康署(HealthCanada)深度抽吸方法,其强调测量每种卷烟产品的最大可能烟气释放量。为了确定采用新的卷烟抽吸实验的ISO方法(深度抽吸方法)后,我们卷烟厂的主导产品,“长白山”系列卷烟的常规烟气指标(焦油、烟碱、CO量)实测值的变化情况,我们将采用加拿大健康署的深度抽吸方法和国际标准方法对我们卷烟厂“长白山”系列卷烟的常规烟气指标进行测定和比较,讨论抽吸方法的改变对我们卷烟厂“长白山”卷烟烟气释放量的影响。

不论最终采用哪种新的抽吸方法,都会导致卷烟烟气释放量测定值的大幅增加。这使我厂低焦油,特别是滤嘴打通风孔的卷烟面临考验,给低焦油卷烟市场带来巨大冲击。因此研究我厂“长白山”牌卷烟的最大烟气释放量究竟是多少?滤嘴通风孔被封闭、抽吸方案的改变对卷烟各项数据的影响到底

有多大?ISO标准修改后我厂的卷烟是否还能称为低焦油卷烟?我厂应该采取什么样的措施来面对?这些都需要我们企业提前了解,及早开始研究,及早做好相应的准备。使“长白山”牌低焦油卷烟在市场上继续拥有强大的市场竞争力。

## 1. 材料,技术指标、检测方法 & 参数设定

根据现有卷烟的规格细支,中支,标准值进行选择。按照GB/T 5606.1—2004 [9] 和 GB/T 16447—2004 [10] 的方法抽样并调节卷烟样品。分别在ISO抽吸模式和深度抽吸模式下,对每个样品进行3次平行测试,各测试项目检测方法见表:

表1 常规指标的检测项目及方法

测试项目	检测方法
焦油	GB/T19609-2004
烟碱	YC/T156-2001
一氧化碳	YC/T30-199

表2 两种抽吸模式下的抽吸参数的设定

抽吸模式	抽吸容量	抽吸频率	抽吸持续时间/S	滤嘴通风口封闭率%
ISO 模式	35±0.3	60±0.5	2±0.02	0
加拿大模式	55±0.5	30±0.5	2±0.02	100

## 2.1 实验设备

RM200A2型20孔道转盘型吸烟机配有COA205一氧化碳测定仪(公司,德国),7890气相色谱仪配有FID检测器、TCD检测器(安捷伦公司,美国),sodimax综合测试台(法国索定),KBF-240恒温恒湿箱(德国binder),xs204电子天平(梅特勒-托利)

## 2.2 实验试剂和实验用品

测定焦油所需要的试剂与材料:

异丙醇(色谱纯,DIKMA公司,美国)

无水乙醇(99.7%,分析纯,北京化工厂)

C17(SIGMA公司,美国)

尼古丁纯品(99.3%,从郑州烟草研究院购买)

蒸馏水

直径44 mm玻璃纤维滤片(WHATMAN公司 英国)

水分测定用气相色谱填充柱（内径2 mm，长度2 m，固定相Chromosorb102，八方公司）

尼古丁测定用气相色谱毛细柱（内径2 mm，长度2 m，固定相100目~80目酸洗的硅烷化担体上涂10%PEG20M+2%KOH，八方公司）。

### 2.3 抽吸方案的制定与样品制备

#### 2.3.1 采用加拿大健康署吸烟标准的抽吸方案的制定

使用RM200A2型20孔道转盘型吸烟机每次抽吸1个规格的样品，

1-20孔道抽吸1种样品，每个孔道抽吸1支卷烟。抽吸容量 $55 \pm 0.3$  mL、抽吸持续时间 $2 \pm 0.05$  s、采用直径90mm玻璃纤维滤片（已在测试大气中调节至少12h）收集20支样品，环境温烟标准的样品制备

将“长白山”牌系列4度 $22 \pm 2$  °C，相对湿度 $60 \pm 5\%$ 。

#### 2.3.2 采用加拿大健康署吸

4个规格的产品（3mg卷烟、8mg卷烟、9mg卷烟、7mg卷烟）用17mm宽的胶纸将打孔的烟的通风孔全部封闭。在相对湿度 $60 \pm 3\%$ 温度 $22 \pm 2$  °C的条件下调节48小时。测定所有烟支的吸阻和通风率确保烟支不漏气。

#### 2.3.3 采用国家吸烟标准

##### 2.3.3.1 采用国家吸烟标准的抽吸方案的制定

使用RM200A2型20孔道转盘型吸烟机每次抽吸1个规格的样品，

1-20孔道抽吸1种样品，每个孔道抽吸1支卷烟。抽吸容量 $35 \pm 0.3$  mL、抽吸持续时间 $2 \pm 0.05$  s，采用直径90 mm玻璃纤维滤片（已在测试大气中调节至少12h）收集20支样品，环境温度 $22 \pm 2$  °C，相对湿度 $60 \pm 5\%$ 。

##### 2.3.3.2 采用国家吸烟标准的样品制备

将4个规格的产品（3mg卷烟、8mg卷烟、9mg卷烟、7mg卷烟）在相对湿度 $60 \pm 3\%$ ，温度 $22 \pm 2$  °C的条件下调节48小时。测定所有烟支的吸阻和通风率。

### 2.4 烟气粒相物的捕集与测定

#### 2.4.1 采用加拿大健康署吸烟标准

将RM200A2型吸烟机调到加拿大健康署抽吸参数。每次抽吸1个牌号样品，20个孔道抽吸1种样品，每个孔道抽吸1支卷烟，将已经在测试大气中调节12 h的滤片放入滤片夹持器中，滤片粗糙的一面面向进入的烟气，合上滤片夹持器，盖上两端的密封帽。并在同一吸烟环境下准备两个水分空白。称量烟气捕集器的质量，精确到0.1 mg，将捕集器装到吸烟机的吸烟孔道上完成吸烟过程，取下捕集器盖好两端密封帽，精确称量到0.1 mg，计算出每支烟的总粒相物质量，用COA205一氧

化碳测定仪测定每个孔道的一氧化碳释放量。

#### 2.4.2 采用国家标准

将RM200A2型吸烟机调到国家标准抽吸参数。每次抽吸1个产品，每个孔道抽吸1支卷烟。将已经在测试大气中调节12 h的滤片放入滤片夹持器中，滤片粗糙的一面面向进入的烟气，合上滤片夹持器，盖上两端的密封帽。并在同一吸烟环境下准备两个水分空白。称量烟气捕集器的质量，精确到0.1 mg，将捕集器装到吸烟机的吸烟孔道上完成吸烟过程，取下捕集器盖好两端密封帽，精确称量到0.1 mg，计算出每支烟的总粒相物质量，用COA205一氧化碳测定仪测定每个孔道的一氧化碳释放量。

### 2.5 水和尼古丁的测定

#### 2.5.1 色谱溶剂的配制

配制含有合适浓度内标物的异丙醇。水分测定内标物为无水乙醇（纯度99.7%），浓度约为5 mL/L。尼古丁测定内标物碳17，浓度约为0.3 mL/L。

#### 2.5.2 水的测定

##### 2.5.2.1 检测水分时气相色谱条件

柱箱温度：170 °C，进样口温度：240 °C，载气：氩气，载气流速：10 mL/min，检测器温度：250 °C，进样体积：1  $\mu$ L

##### 2.5.2.2 制定水分定量工作曲线

分别取2  $\mu$ L,8  $\mu$ L,16  $\mu$ L,20  $\mu$ L,25  $\mu$ L蒸馏水加入50 mL色谱溶剂中，配制成不同浓度的水标准溶液，按2.5.2.1条件进行测定，计算线性回归方程，以此计算样品中水分含量。

##### 2.5.2.3 水的测定

将每个捕集过的滤片对折放入到1个在测试环境中平衡12小时以上的100 mL锥形瓶中，并用约35 mg的脱脂棉擦拭捕集器前盖内壁，放入锥形瓶中，移入20 mL色谱溶剂萃取，振荡30分钟，用移液器将上述萃取液移入2 mL色谱瓶。按2.5.2.1条件进行测定，计算出每支烟烟气粒相物中的水分。

#### 2.5.3 尼古丁的测定

##### 2.5.3.1 检测尼古丁时气相色谱条件

柱箱温度：170 °C，进样口温度：240 °C，载气：氮气，载气流速：50 mL/min，检测器温度：250 °C，进样体积：1  $\mu$ L

##### 2.5.3.2 尼古丁母液的配制

准确称量250 mg尼古丁纯品，精确至0.01 mg，置于250 mL容量瓶中，用异丙醇标定到刻度，尼古丁母液浓度

约1 mg/mL

2.5.3.3 制定尼古丁定量工作曲线

分别取1 mL, 2 mL, 3 mL, 4 mL, 5 mL, 6 mL尼古丁母液(约1 mg/mL), 加入20 mL色谱溶剂中, 配制成不同浓度的尼古丁标准溶液, 按2.5.3.1条件进行测定, 计算线性回归方程, 以此计算样品中尼古丁含量。

2.5.3.4 尼古丁的测定

将每个捕集过的滤片对折放入到1个在测试环境中平衡12小时以上的100 mL锥形瓶中, 并用约35 mg的脱脂棉擦拭捕集器前盖内壁, 放入锥形瓶中, 移入20 mL色谱溶剂萃取, 振荡30分钟, 用移液器将上述萃取液移入2 mL色谱瓶。按2.5.3.1条件进行测定, 计算出每支卷烟气相物中的尼古丁量。

2.6 焦油量的计算

焦油Tar, 以每支烟的毫克数表示, 按式(1)计算:  
 $Tar = M\ TPM - M\ H_2O - M\ NIC \dots\dots\dots (1)$

式中:

- M TPM ——总粒相物的质量, 单位为毫克 (mg);
- M H<sub>2</sub>O ——总粒相物中的水分质量, 单位为毫克 (mg);
- M NIC ——总粒相物中的烟碱质量, 单位为毫克 (mg)。

3、结果和讨论

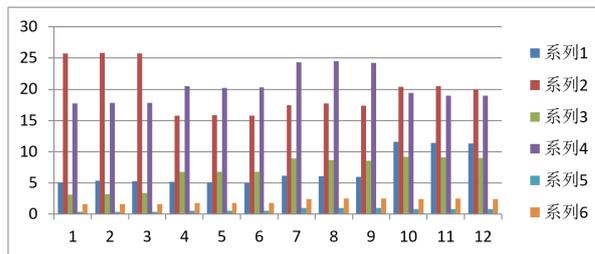
3.1 检测结果

本次实验使用了4个规格的产品通过采用加拿大深度抽取方法和国标抽取方法, 我们得到了4种卷烟主流烟气各项检测指标测定结果(表3-表4)。

表3 两种抽吸模式下不同卷烟主流烟气CO、焦油、烟碱释放量

样品	CO/(mg·支 <sup>-1</sup> )		焦油(mg·支 <sup>-1</sup> )		烟碱(mg·支 <sup>-1</sup> )	
	ISO	加拿大	ISO	加拿大	ISO	加拿大
3mg 卷烟	5.1	25.7	3.1	17.7	0.4	1.6
	5.4	25.8	3.2	17.8	0.4	1.6
	5.3	25.7	3.4	17.8	0.4	1.6
7mg 卷烟	5.2	15.8	6.8	20.5	0.6	1.8
	5.1	15.9	6.8	20.2	0.6	1.8
	5.0	15.8	6.8	20.3	0.6	1.8
9mg 卷烟	6.2	17.5	8.9	24.3	1.0	2.4
	6.1	17.7	8.7	24.5	1.0	2.5
	6.0	17.4	8.6	24.2	1.0	2.5
8mg 卷烟	11.6	20.4	9.2	19.4	0.8	2.4
	11.4	20.5	9.1	19.0	0.8	2.5
	11.3	20.0	9.0	19.0	0.8	2.4

图1 两种抽吸模式下卷烟烟气中CO、焦油、烟碱释放量条形图



系列1—co (iso) 系列2—co (加拿大); 系列3—焦油 (iso) 系列4--焦油 (加拿大);

系列5—烟碱 (iso) 系列6—烟碱 (加拿大);

小结:从化学成分释放量增大倍数上可以看出

3mg卷烟焦油释放量平均增大5.8倍, 烟碱释放量平均增大4.0倍, 一氧化碳释放量平均增大5.0倍

7mg卷烟焦油释放量平均增大3.0倍, 烟碱释放量平均增大3.0倍, 一氧化碳释放量平均增大3.1倍

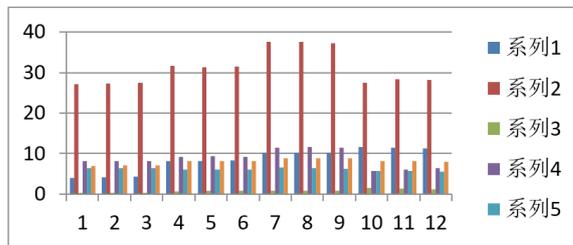
9mg卷烟焦油释放量平均增大2.8倍, 烟碱释放量平均增大2.4倍, 一氧化碳平均增大2.9倍;

8mg卷烟焦油释放量平均增大2.1倍, 烟碱释放量平均增,3倍, 一氧化碳平均增大1.8倍

表4 两种抽吸模式下不同卷烟主流烟气中TPM、水分、口数释放量

样品	TPM (mg·支 <sup>-1</sup> )		水分 (mg·支 <sup>-1</sup> )		口数	
	ISO	加拿大	ISO	加拿大	ISO	加拿大
3mg 卷烟	4.0	27.2	0.31	8.2	6.4	7.0
	4.1	27.3	0.31	8.1	6.5	7.1
	4.3	27.5	0.32	8.1	6.4	7.1
7mg 卷烟	8.2	31.6	0.7	9.3	6.1	8.1
	8.2	31.4	0.8	9.4	6.1	8.1
	8.3	31.5	0.8	9.3	6.1	8.2
9mg 卷烟	10.3	37.6	0.9	11.4	6.6	8.9
	10.2	38.5	0.9	11.6	6.5	8.9
	10.2	37.3	0.8	11.4	6.3	8.9
8mg 卷烟	11.6	27.5	1.5	5.8	5.8	8.2
	11.4	28.4	1.4	6.0	5.7	8.1
	11.3	28.2	1.2	6.4	5.5	8.0

图2 两种抽吸模式下卷烟烟气中TPM、水分、口数释放量条形图



系列1—TPM (iso) 系列2—TPM (加拿大)；系列3—水分 (iso) 系列4—水分 (加拿大)；系列5—口数 (iso) 系列6—口数 (加拿大)；

小结：

从化学成分释放量数据变化上可以看出：

3mg卷烟总粒相物释放量增大6.5倍，水分增大26倍，口数增大1.1倍

7mg卷烟总粒相物释放量平均增大3.9倍，水分平均增大12.4倍，口数平均增大1.3倍（基本不变）

9mg卷烟总粒相物平均增大3.7倍，水分平均增大13.5倍，口数平均增大1.4倍

8mg卷烟总粒相物释放量平均增大2.4倍，水分平均增大4.5倍，口数平均增大1.4倍

### 3.2 数据分析

#### 3.2.1 实验结果的描述性分析

从以上表2和表3烟气释放量数据变化可以得出：深度抽吸模式下，主流烟气中焦油、烟碱和水分,CO,TPM释放量明显高于 ISO 抽吸模式下的释放量，其原因主要是深度抽吸模式与 ISO 抽吸模式相比，采用更大的抽吸容量、更小的抽吸频率且其滤嘴通风孔被完全封闭，完全抵消了滤嘴通风的降焦效果，因此主流烟气中焦油、烟碱和水分CO,TPM释放量明显高于 ISO 抽吸模式下的释放量。

深度抽吸模式下，主流烟气中有通风率的卷烟水分增加的倍数非常大,可见滤嘴通风率和抽吸容量是影响主流烟气中水分释放量的重要因素.封闭通风孔的卷烟，其常规烟气成分释放量较没有通风孔的卷烟采用以上两种方法测得的结果差异明显增大。

实验过程中也呈现一个明显的规律，在卷烟的实际抽吸长度一致的情况下，封闭滤嘴通风孔对高透气度超低焦油卷烟烟气释放量测定结果的影响要大于较低透气度低焦油卷烟的影响，更大于没有采用滤嘴打孔技术的卷烟。

#### 3.2.2 两种抽吸模式下焦油释放量的相关性分析

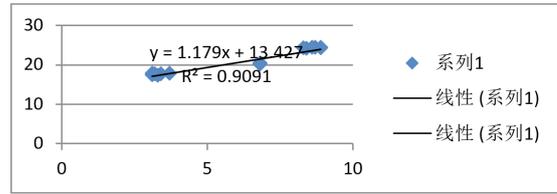
采用datahoop 2.0平台对两种抽吸模式下焦油释放量进行回归分析

得到以下分析结果:注（只对有通风稀释率的卷烟）

R 方	调整 R 方	F	P 值
0.909121434	0.903441523	160.0591158	9.53008E-10

模型	系数	p 值 (t 检验)	置信区间下限	置信区间上限	预测区间下限	预测区间上限
iso	1.1789992	9.53008E-10	0.981443408	1.376554989	-7.426306262	9.784304658
常数项	13.4266939	2.63373E-13	12.11744888	14.73593891	4.821388436	22.03199936

以ISO测得的焦油含量为x，以深度抽吸模式测得的焦油含量为Y绘制散点图如下



用相关系数矩阵分析可得相关系数为0.9534。从图和相关系数都可以看出ISO测得的焦油含量和深度抽吸模式测得的焦油含量有比较强的正相关关系。

#### 3.2.3 两种抽吸模式下焦油释放量的预测性分析

以ISO测得的焦油含量为自变量x，以深度抽吸模式测得的焦油含量为因变量Y,做线性回归预测分析，得到回归方程如下：

$$Y=1.18X+13.43$$

回归系数为1.18表示ISO测得的焦油释放量每增加一个单位，深度抽吸模式下测得的焦油释放量大致增加1.18个单位，ISO模式下测得的焦油释放量对深度抽吸模式下测得的焦油释放量的影响是正向的，ISO模式下测得的焦油释放量越高，深度抽吸模式下测得的焦油释放量就越高。

可决系数R方为0.9091，说明模型拟合效果很好

T检验和F检验的P值都小于0.05，线性关系显著，此预测方程成立。

### 4、结论

本项研究采用深度抽吸和ISO两种方法，测定了我厂“长白山”系列卷烟烟气指标的释放量，初步探讨了采用深度抽吸方法对我厂“长白山”卷烟烟气指标释放量测定结果的影响。实验结果表明：采用深度抽吸方法测得的“长白山”系列卷烟烟气指标释放量都明显增加，特别是采用通风稀释技术的卷烟，此种现象更加显著。

建议：滤嘴通风率可由卷烟设计和卷烟材料控制，而抽吸容量和抽吸频率是由消费者的抽吸习惯控制，随着卷烟滤嘴通风度的增加，采用两种方法测得的烟气常规化学成分的释放量差异也显示出增大得趋势。改变滤嘴通风率是目前烟草行业开发低焦油低危害卷烟重要技术手段之一，在加拿大深度抽吸模式下，卷烟的滤嘴通风孔被完全封闭，完全抵消了滤嘴通风的减害降焦效果，导致其作为开发低焦油低危害卷烟的技术手段失效,由此可知，在深度抽吸模式下，迫切需要开发新的减害降焦技术手段，应对深度抽吸模式对烟草行业低危害卷烟的开发带来的挑战。

#### 参考文献：

[1] 王芳，温东奇，陈再根，孔浩辉，李栋。深度吸烟对卷烟焦油、烟碱、和CO释放量测定结果的影响。《烟草科技》，2006年，第3期，总第244期

[2] 用常规分析用吸烟机测定总粒相物和焦油，GB/T19609-2004。

## / Python提速的10个小技巧，你值得拥有 /

来源 / 数据说 编辑 / 协会会员处 李苗苗 日期 / 2020-09



一些小提示和小技巧可能是非常有用的，特别是在编程领域。有时候使用一点点黑客技术，既可以节省时间，还可能挽救“生命”。

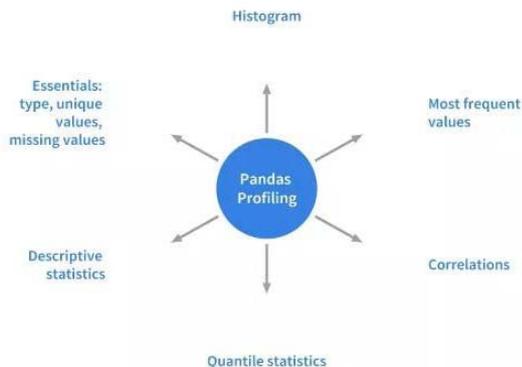
一个小小的快捷方式或附加组件有时真是天赐之物，并且可以成为真正的生产力助推器。所以，这里有一些小提示和小技巧，有些可能是新的，但我相信在下一个数据分析项目中会让你非常方便。

### Pandas中数据框数据的Profiling过程

Profiling（分析器）是一个帮助我们理解数据的过程，而Pandas Profiling是一个Python包，它可以简单快速地对Pandas的数据框数据进行探索性数据分析。

Pandas中df.describe()和df.info()函数可以实现EDA过程第一步。但是，它们只提供了对数据非常基本的概述，对于大型数据集没有太大帮助。而Pandas中的Profiling功能简单通过一行代码就能显示大量信息，且在交互式HTML报告中也是如此。

对于给定的数据集，Pandas中的profiling包计算了以下统计信息：



由Pandas Profiling包计算出的统计信息包括直方图、众数、相关系数、分位数、描述统计量、其他信息——类型、单一变量值、缺失值等。

### 安装

用pip安装或者用conda安装

```
1 pip install pandas-profiling
```

```
2 conda install -c anaconda pandas-profiling
```

### 用法

下面代码是用很久以前的泰坦尼克数据集来演示多功能Python分析器的结果。

```
1 #importing the necessary packages
```

```
2 import pandas as pd
```

```
3 import pandas_profiling
```

```
4 df = pd.read_csv('titanic/train.csv')
```

```
5 pandas_profiling.ProfileReport(df)
```

一行代码就能实现在Jupyter Notebook中显示完整的数据分析报告，该报告非常详细，且包含了必要的图表信息。

```
import pandas_profiling
pandas_profiling.ProfileReport(df)
```

还可以使用以下代码将报告导出到交互式HTML文件中。

```
1 profile = pandas_profiling.ProfileReport(df)
```

```
2 profile.to_file(outputfile="Titanic data profiling.html")
```

## Overview

### Dataset info

Number of variables	12
Number of observations	891
Total Missing (%)	8.1%
Total size in memory	83.6 KiB
Average record size in memory	96.1 B

### Variables types

Numeric	6
Categorical	4
Boolean	1
Date	0
Text (Unique)	1
Rejected	0
Unsupported	0

### Warnings

Age has 177 / 19.9% missing values Missing

Cabin has 687 / 77.1% missing values Missing

Cabin has a high cardinality: 148 distinct values Warning

Fare has 15 / 1.7% zeros Zeroes

Parch has 678 / 76.1% zeros Zeroes

SibSp has 608 / 68.2% zeros Zeroes

Ticket has a high cardinality: 681 distinct values Warning

### Pandas实现交互式作图

Pandas有一个内置的.plot () 函数作为DataFrame类的一部分。但是，使用此功能呈现的可视化不是交互式的，这使得它没那么吸引人。同样，使用pandas.DataFrame.plot () 函数绘制图表也不能实现交互。如果我们需要在不对代码进行重大修改的情况下用Pandas绘制交互式图表怎么办呢？这个时候就可以用Cufflinks库来实现。

Cufflinks库可以将有强大功能的plotly和拥有灵活性的pandas结合在一起，非常便于绘图。下面就来看看在pandas中如何安装和使用Cufflinks库。

#### 安装

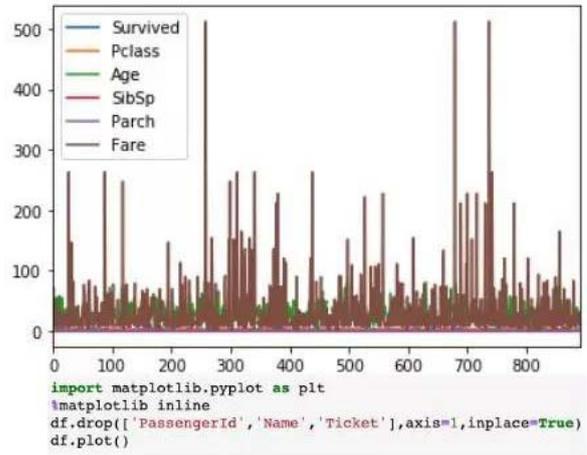
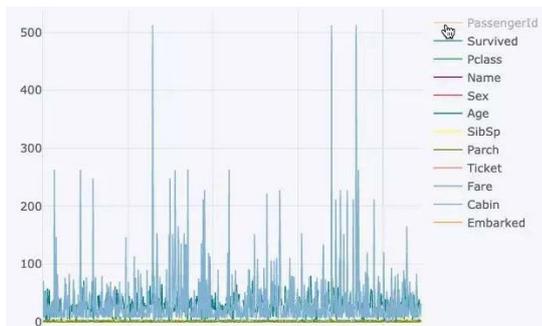
```
1 pip install plotly
2 # Plotly is a pre-requisite before installing cufflinks
3 pip install cufflinks
```

#### 用法

```
1 #importing Pandas
2import pandas as pd
3 #importing plotly and cufflinks in offline mode
4 import cufflinks as cf
5 import plotly.offline
6 cf.go offline()
7 cf.set config file(offline=False, world_readable=True)
```

是时候展示泰坦尼克号数据集的魔力了。

```
1 df.iplot()
```



1 df.iplot() vs df.plot()

右侧的可视化显示了静态图表，而左侧图表是交互式的，更详细，并且所有这些在语法上都没有任何重大更改。

#### Magic命令

Magic命令是Jupyter notebook中的一组便捷功能，旨在解决标准数据分析中的一些常见问题。使用命令%ismagic可以看到所有的可用命令。

```
%ismagic

Available line magics:
%alias %alias_magic %autoawait %autocall %automagic %autosave %bookmark %cat %cd %cl
ear %colors %conda %config %connect_info %cp %debug %dhist %dirs %doctest_mode %ed
%edit %env %gui %hist %history %killbgscripts %ldir %less %lf %lk %ll %load %load
_ext %loadpy %logoff %logon %logstart %logstate %logstop %ls %ls_magic %lx %macro %
magic %man %matplotlib %mkdir %more %mv %notebook %page %pastebin %pdb %pdef %pdoc
%profile %pinfo %ppinfo2 %pip %popd %pprint %precision %prun %psearch %psource %pushd
%pwd %pycat %pylab %qtconsole %quickref %recall %rehashx %reload_ext %rep %rerun %r
eset %reset_selective %rm %rmdir %run %save %sc %set_env %store %sx %system %tb %
time %timeit %unalias %unload_ext %who %who_ls %whos %xdel %xmode

Available cell magics:
%%! %%HTML %%SVG %%bash %%capture %%debug %%file %%html %%javascript %%js %%latex
%%markdown %%perl %%prun %%ppyy %%python %%python2 %%python3 %%ruby %%script %%sh %%
svg %%sx %%system %%time %%timeit %%writefile

Automagic is ON, % prefix IS NOT needed for line magics.
```

#### 所有可用的Magic命令列表

Magic命令有两种：行magic命令（line magics），以单个%字符为前缀，在单行输入操作；单元magic命令（cell magics），以双%%字符为前缀，可以在多行输入操作。如果设置为1，则不用键入%即可调用Magic函数。



接下来看一些在常见数据分析任务中可能用到的命令：

% pastebin

%pastebin将代码上传到Pastebin并返回url。Pastebin是一个在线内容托管服务，可以存储纯文本，如源代码片段，然后通过url可以与其他人共享。事实上，Github gist也类似于pastebin，只是有版本控制。

在file.py文件中写一个包含以下内容的python脚本，并试着运行看看结果。

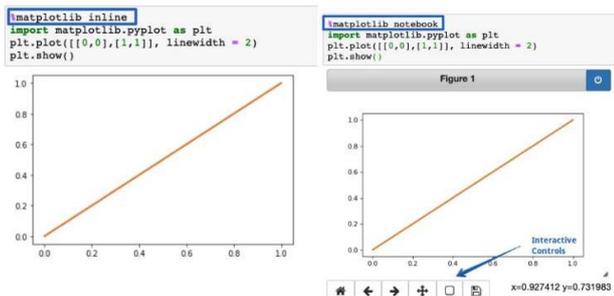
```
1 #file.py
2 def foo(x):
3 return x
```

在Jupyter Notebook中使用%pastebin生成一个pastebin url。



%matplotlib notebook

函数用于在Jupyter notebook中呈现静态matplotlib图。用notebook替换inline，可以轻松获得可缩放和可调整大小的绘图。但记得这个函数要在导入matplotlib库之前调用。



%run

用%run函数在notebook中运行一个python脚本试试。

```
1 %run file.py
2 %%writefile
```

%% writefile是将单元格内容写入文件中。以下代码将脚本写入名为foo.py的文件并保存在当前目录中。

```
%%writefile foo.py
def foo(x):
return x
```

Writing foo.py

%latex

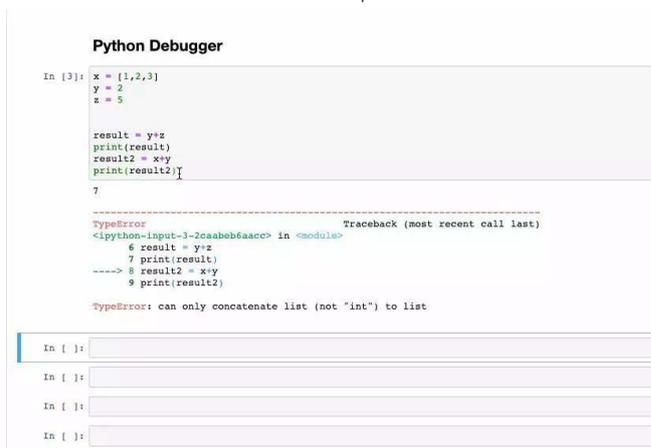
%latex函数将单元格内容以LaTeX形式呈现。此函数对于在单元格中编写数学公式和方程很有用。

```
%%latex
\begin{align}
a = \frac{1}{2} && b = \frac{1}{3} \\
\end{align}
```

$$a = \frac{1}{2} \quad b = \frac{1}{3}$$

## 查找并解决错误

交互式调试器也是一个神奇的功能，我把它单独定义了一类。如果在运行代码单元时出现异常，请在新行中键入% debug并运行它。这将打开一个交互式调试环境，它能直接定位到发生异常的位置。还可以检查程序中分配的变量值，并在此处执行操作。退出调试器单击q即可。



## Printing也有小技巧

如果您想生成美观的数据结构，pprint是首选。它在打印字典数据或JSON数据时特别有用。接下来看一个使用print和pprint来显示输出的示例。

```
# With Print
employee_records = {'Emp ID': '101', 'Emp Name': 'Tom',
                    'Project IDs': {'P1':1308, 'P2': 'A104', 'P4':2}}
print(employee_records)
{'Emp ID': '101', 'Emp Name': 'Tom', 'Project IDs': {'P1': 1308, 'P2': 'A104', 'P4': 2}}
```

```
# With Pretty Print
import pprint
employee_records = {'Emp ID': '101', 'Emp Name': 'Tom',
                    'Project IDs': {'P1':1308, 'P2': 'A104', 'P4':2}}
pprint.pprint(employee_records,width=-1)
{'Emp ID': '101',
 'Emp Name': 'Tom',
 'Project IDs': {'P1': 1308,
                 'P2': 'A104',
                 'P4': 2}}
```

## 让你的笔记脱颖而出

我们可以在您的Jupyter notebook中使用警示框/注释框来突出显示重要内容或其他需要突出的内容。注释的颜色取决于指定的警报类型。只需在需要突出显示的单元格中添加以下任一代码或所有代码即可。



### 蓝色警示框：信息提示

```
1 <div class="alert alert-block alert-info">
2 <b>Tip:</b> Use blue boxes (alert-info) for tips and notes.
3 If it's a note, you don't have to include the word "Note".
4 </div>
```

Tip:Info: The Blue boxes are used for tips and notes.

### 黄色警示框：警告

```
1 <div class="alert alert-block alert-warning">
2 <b>Example:</b> Yellow Boxes are generally used to include
3 additional examples or mathematical formulas.
4 </div>
```

Yellow Boxes are generally used to include additional examples or mathematical formulas.

### 绿色警示框：成功

```
1 <div class="alert alert-block alert-success">
2 Use green box only when necessary like to display links to
3 related content.
4 </div>
```

Use green box only when necessary like to display links to related content.

### 红色警示框：高危

```
1 <div class="alert alert-block alert-danger">
2 It is good to avoid red boxes but can be used to alert users
3 to not delete some important part of code etc.
4 </div>
```

It is good to avoid red boxes but can be used to alert users to not delete some important part of code etc.

### 打印单元格所有代码的输出结果

假如有一个Jupyter Notebook的单元格，其中包含以下代码行：

```
1 In [1]: 10+5
2      11+6
3 Out [1]: 17
```

单元格的正常属性是只打印最后一个输出，而对于其他输出，我们需要添加print()函数。然而通过在notebook顶部添加以下代码段可以一次打印所有输出。

添加代码后所有的输出结果就会一个接一个地打印出来。

```
1 In [1]: 10+5
2      11+6
3      12+7
4 Out [1]: 15
5 Out [1]: 17
6 Out [1]: 19
```

恢复原始设置：

```
1 InteractiveShell.ast_node_interactivity = "last expr"
```

使用'!'选项运行python脚本

从命令行运行python脚本的典型方法是：  
python hello.py

但是，如果在运行相同的脚本时添加-i，例如python -i hello.py，就能提供更多优势。接下来看看结果如何。

首先，即使程序结束，python也不会退出解释器。因此，我们可以检查变量的值和程序中定义的函数的正确性。



其次，我们可以轻松地调用python调试器，因为我们仍然在解释器中：

```
1 import pdb
2 pdb.pm()
```

这能定位异常发生的位置，然后我们可以处理异常代码。

### 自动评论代码

Ctrl / Cmd + / 自动注释单元格中的选定行，再次命中组合将取消注释相同的代码行。

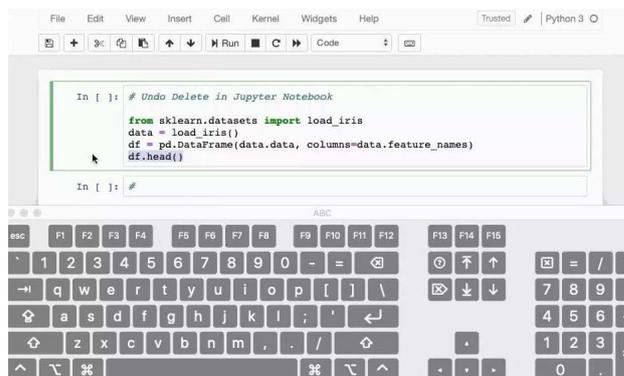
```
In [ ]: from sklearn.linear_model import LinearRegression
        model = LinearRegression()
        model.fit(dt_df[:,["loan_amnt", "installment"]], dt_df[:, "int_rate"])
        model.coef_
```

### 删除容易恢复难

你有没有意外删除过Jupyter notebook中的单元格？如果答案是肯定的，那么可以掌握这个撤消删除操作的快捷方式。

如果您删除了单元格的内容，可以通过按CTRL / CMD + Z轻松恢复它。

如果需要恢复整个已删除的单元格，请按ESC + Z或EDIT>撤消删除单元格。



### 结论

在本文中，我列出了使用Python和Jupyter notebook时收集的一些小提示。我相信它们会对你有用，能让你有所收获，从而实现轻松编码！

# / 阿里巴巴大牛是怎么写数据分析报告的? /

来源 / 数据说 编辑 / 协会会员处 李苗苗 日期 / 2020-09



在当今企业纷纷推动数字化运营的背景下，“No Data, No BB”成了职场人的口头禅。做一份好的数据分析报告，大到成为能否帮助企业做出正确的商业决策，小到成为能否说服老板获取业务资源的关键因素。今天数据君和大家一起来学习一位阿里巴巴大牛数据分析师撰写分析报告的经验之谈：

## 1. 表达主题决定了我们的图表形式

决定分析报告图表形式的并不是拥有的数据是什么，而是你所需要表达的主题是什么。图1和图2是根据一份相同的数据，展现的2个不同的图表：

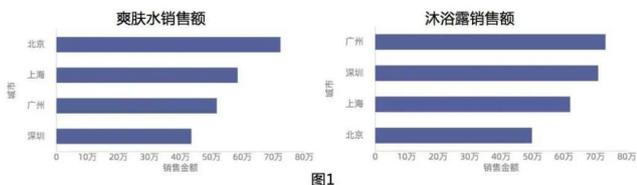


图1

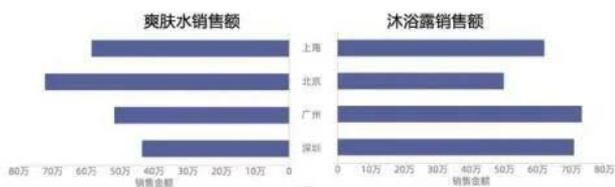


图2

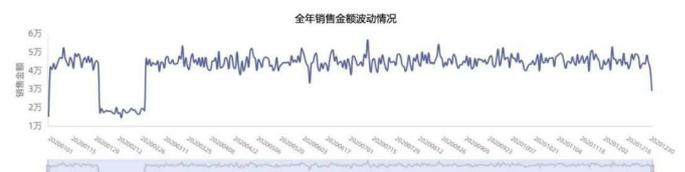
上图可以发现对于相同的数据，因为我们所需表达的主题的差异，也将呈现完全不同的展现方式。

图1表达的主题是爽肤水和沐浴露两个品类在不同城市的销量排名，图2主要表达的主题是在相同城市在两个不同的品类的销量差异。因此在下笔做分析报告之前，先仔细想好自己要表达的主题是什么。

## 2. 不要放弃“标题”这个绝佳位置

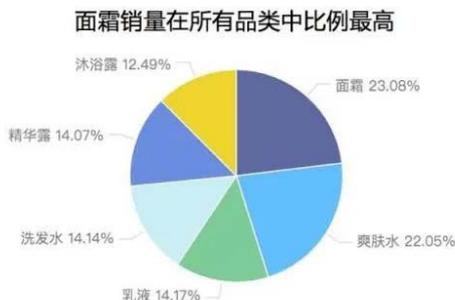
有些图表的标题就和猜谜一样，例如：公司销售趋势、分公司销售分布情况。完全没有指出图表的重点，公司销售趋势是怎么样的？分公司销售分布又是如何？我们需要强调的重点应该放在图表最前面，减少听众误解的可能性，并让他们的注意力集中到我们所想强调的数据上。

如下图，这张图到底是为了表达全量销售金额没有明显的增长呢？还是为了表达2月份销售金额断崖式下跌？还是为了表达其它什么主题呢？请在标题上明确的告诉听众。



### 3. 处理“成分对比”关系

成分对比主要体现在一个整体的每个部分的百分比的对比，常常出现“份额”、“百分比”等词汇，并使用饼图来展现。



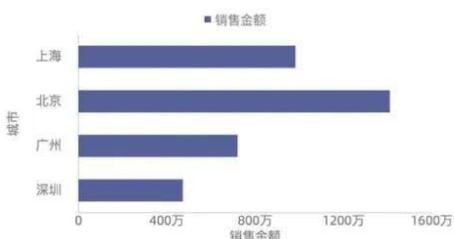
饼图在使用中建议不超过6个部分，如果超过6个部分，可以把剩余部分归类到“其它”项中。另外由于人们看数据习惯顺时针看数据，因此可把最重要部分放到12点位置，并用对比度强烈的颜色突出显示。

饼图主要在标识单一整体各部分比例，如果需要比较两个整体的成分时，重点考虑柱状图（图3）。因为如果使用饼图（图4）会导致读者视线需要在不同图表间来回移动。



### 4. 处理“项目间对比”关系

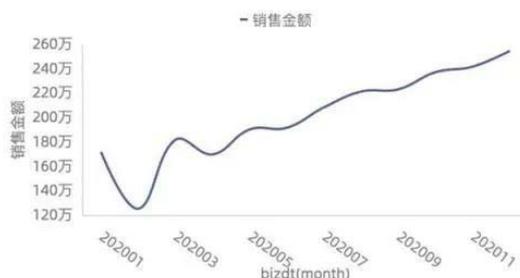
项目间对比主要是比较不同项目间的情况，常常出现“排名”、“大小”等词汇。项目间对比通常使用条形图来展示。



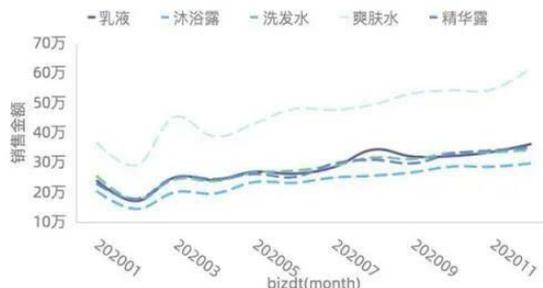
我们对于条形图的顺序需要深思熟虑，如果天然是有顺序的则按自然的顺序，例如人生阶段婴儿、少年、青年、中年、老年。但是如果如果没有这种天然顺序，需要考虑什么顺序对于我们数据主题是最有意义的，并根据我们需要突出的主题，选择条形图的排序方式。

### 5. 处理“时间序列对比”关系

时间序列对比关心的是随时间变化的对比，常常出现“变化”、“增长”、“下降”等词汇。时间序列对比通常使用柱状图或折线图来展示，如果时间点不多时可以使用柱状图，如果时间点是很长一段时间范围使用折线图更为合适。



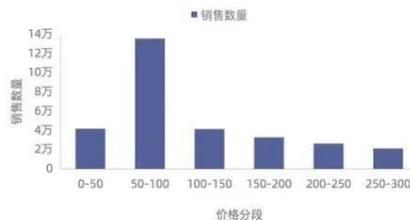
对于折线图，趋势线一定要比背景线粗。当存在同一张折线图存在多条折线时，需要将最关注的线加粗加亮。但是当出现非常多折线时，我们的折线图就会呈现出“方便面食”图表，往往导致图表混乱。如下图：



### 6. 处理“频率分布对比”关系

频率分布对比表现的是数据分布范围情况，常常出现“范围”、“密度”、“分布”等词汇。频率分布对比通常使用柱状图或折线图来展示，当比较范围数量较多时可使用折线图，较少时可通过柱状图。

#### 大多数销售集中在50-100价格区间



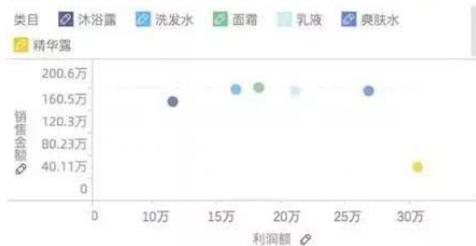
频率分布的范围大小非常重要，既不能太大也不能太

小，建议5到20个分组。不同分组的大小应相同，否则会造成数据扭曲。

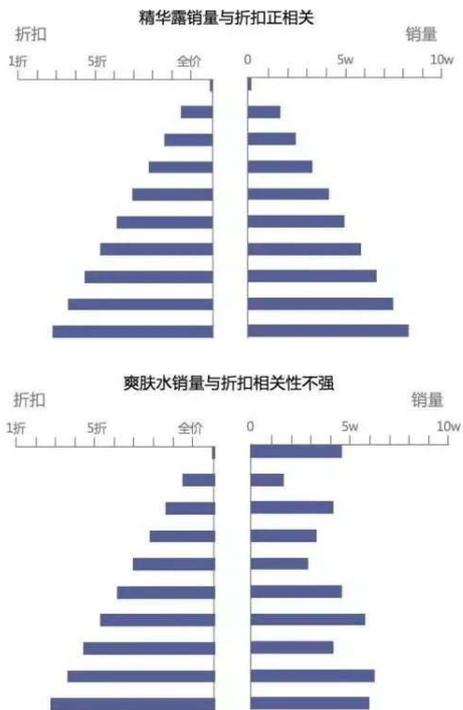
## 7. 处理“相关性对比”关系

相关性对比表现的是不同变量之间的关系，常常出现“与XX有关”、“随XX增长”等词汇，通常使用散点图或双条形图来展示。

精华露相比其它品类，呈现销售金额虽然不高，但利润高的特征



在双条形图中，我们将独立变量按顺序排在左边，而把对比值放在右边，如果期望模式与实际模式一致时，右边的条形图就会变成左边的条形图的镜像，如下面左图。当关系不符合预期时，两组条形图则会发生偏离，如下图：



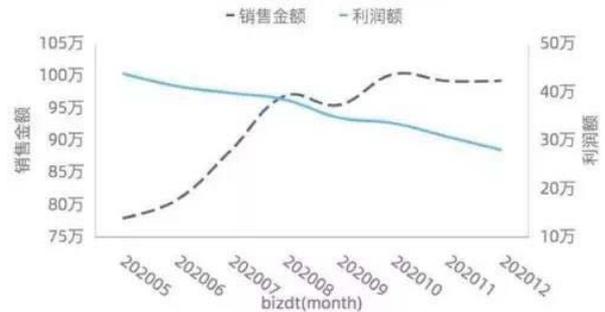
## 8. 处理“多重对比”关系

对比可能不仅仅是一种对比关系，有可能是结合上文所说的成分、项目间、时间序列、频率分布、相关性对比中的一种或多种组合而成，我们称之为“多重对比”。

例如“销售额在过去10年内稳步增长，但利润却没有同步增长”，这案例第一部分“销售额在过去10年内稳步增

长”属于“时间序列对比”，第二部分“但利润却没有同步增长”属于项目间对比。对于这种情况，我们需要确定哪一种关系是主要的，哪一种是次要的。案例中，第一部分随时间变化是主要的，而销售额与利润项目间对比是次要的，因此最好选择以时间变化对比的折线图，并为每一个项目画一条支线的方式实现，如下图：

销售额逐月上升，但利润却在持续下降

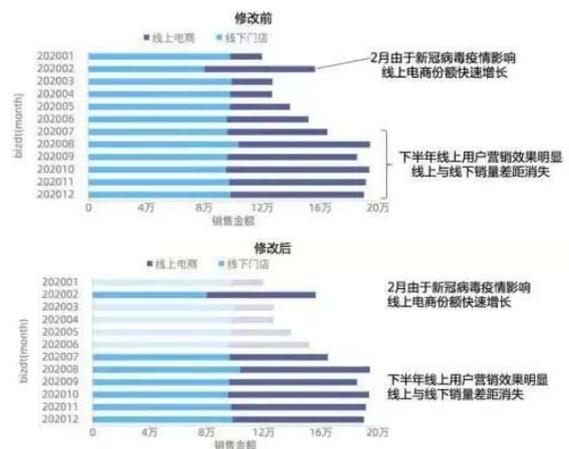


## 9. 不仅仅只有图表

当然，对于数据分析报告，可定量信息我们可以通过图表构建，但是对于不可定量的信息，我们可以通过一些几何图形形成概念性视觉图像，或通过一些日常事务作视觉比喻。例如说到目标，可以形成高山这样的视觉图像。这需要发挥自己的想象力，平时多留意写优秀的海报广告，在生活中寻找灵感，在报告中能够做到与听众产生共鸣的效果。

## 10. 消除认知负荷

听众在接受我们分析报告信息时，需要消耗脑力去学习新知识，脑力是有限的，因此需要消除听众无关紧要的脑力消耗。造成无关紧要脑力消耗最大的问题就是“杂乱”，因此消除杂乱是数据分析报告需要重点关注的。通过将文字从原来居中看齐调为左对齐，进行相关的无关数据的淡化的处理，能减少听众的认知负荷，把关注点转移到我们的重点上。



通过消除一些干扰，能突出我们需要表达的重点。同时要突出我们需要吸引听众实现的地方。在文字中可通过加

粗、颜色、斜体、大小、空间隔离、下划线等手段突出文字关键词。在图表中主要通过颜色、大小突出需要强调的内容。在使用颜色时需慎重选择，不能在一张图中有太多颜色，造成视觉干扰；可以使用颜色的不同饱和度来强调数据；根据分析报告背景，可选择对于的互补色来做内容的突出强调。

## 11. 讲好分析报告

分析报告做好了，还需要以更好的语言表达方式呈现给听众。

可以把分析报告当做一个故事来说，能更加引人入胜。首先对整个背景做设定，接着介绍什么因素上下文驱动情节的演进，举例说明发生了什么样的冲突，基于这些冲突有哪些假设，再基于这些假设如何做数据验证，最后通过什么方式带来什么解决方案。为了保证整个分析报告的逻辑清晰，可以构建类似金字塔的逻辑结构，以某一个中心论点为塔尖，在其以下分支出不同论点的数据分析支撑，让听众对我们的分析报告有个清晰的逻辑结构。

# / 数据分析实战案例：渠道落地页效果分析 /

来源 / 犀数院 编辑 / 协会会员处 李苗苗 日期 / 2020-09

拥有了那么多数据分析相关理论知识，那么应该如何应用呢？今天就来和小编一起进行一次数据分析实战：渠道落地页效果分析。今天给大家分享一个渠道落地页数据分析的实战案例。

## 一、背景和目的

通过对跟渠道落地页数据的分析，优化用户路径，提高各环节的转化率。

## 二、明确问题

根据已经跑了两周的数据，对其投放效果进行分析，寻找现状的问题并产出对该课程投放的优化解决方案。

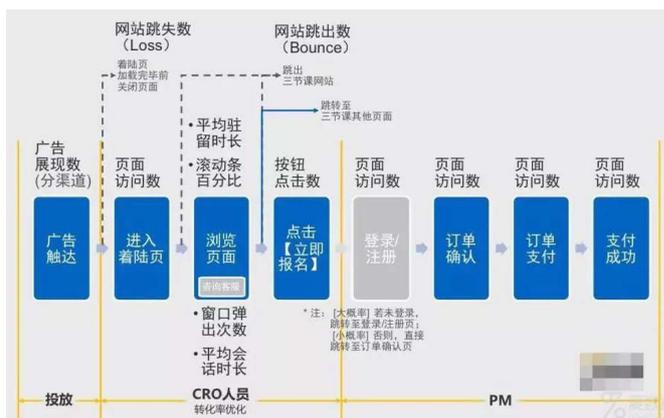
- 点击报名率=注册页UV/着陆页UV
- 注册转化率=订单页UV/注册页UV
- 订单转化率=支付成功页UV/订单页UV

## 三、梳理接触点，形成用户体验路径

用户点击广告后，会直接跳到该着陆页；点击「立即报名」按钮，跳转注册/登录页，如已登录会直接创建订单，点击支付订单按钮，跳转支付页，进行付款。



## 四、确认每一步的衡量指标，并建立漏斗模型



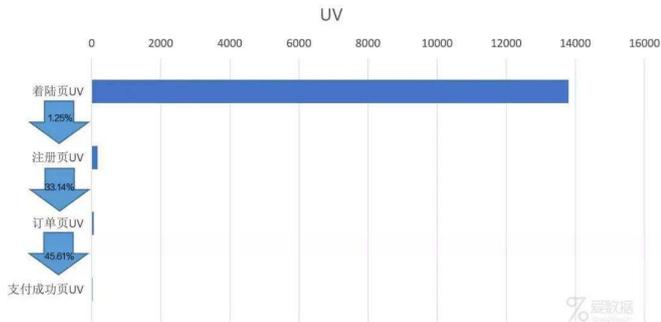
从上图可以看出我们主要需要观察的核心指标是着陆页的uv、注册页的uv、订单确认页的uv、以及支付成功页的uv。

根据已有数据四个渠道的两周的数据进行整理得出如下表格内容：

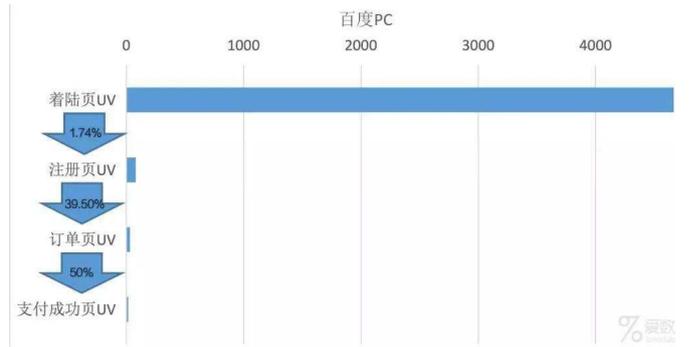
各渠道平均值	百度PC	百度移动	搜狗	360搜索	合计
着陆页UV	4666	5091	2242	1802	13802
注册页UV	81	44	23	24	172
订单页UV	32	10	8	7	57
支付成功页UV	16	5	3	2	26
点击报名率	1.74%	0.86%	1.03%	1.33%	1.25%
注册转化率	39.50%	22.72%	34.78%	29.17%	33.14%
订单转化率	50.00%	50.00%	37.50%	28.57%	45.61%
所有用户在着陆页的平均停留时间 (单位: s)	172	167	101	124	141
页面跳出率	50.77%	54.53%	58.86%	66.00%	57.54%
咨询的占比	22.83%	18.79%	12.68%	9.26%	15.89%
感兴趣值比重	19.96	17.10	7.57	7.57	13.05



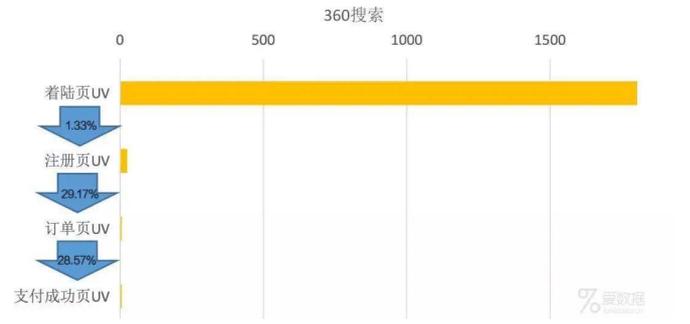
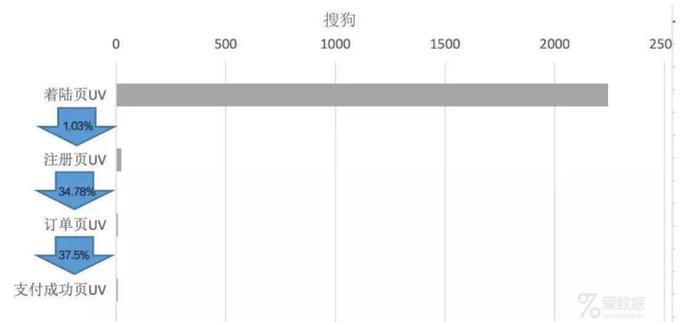
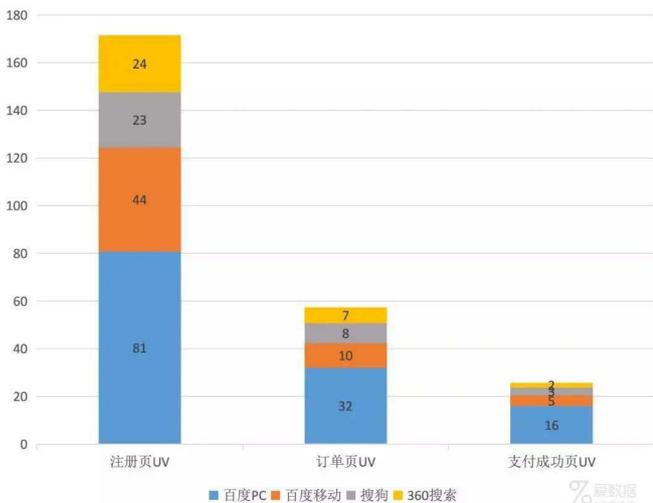
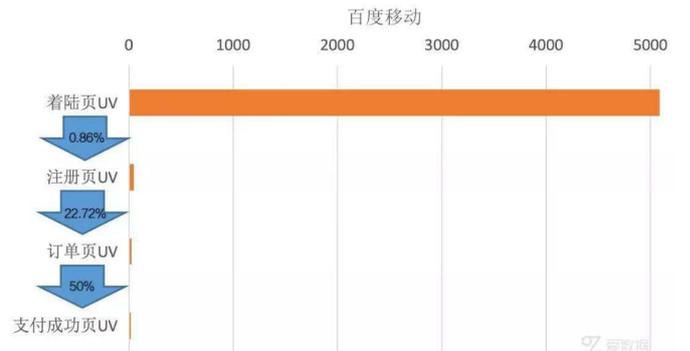
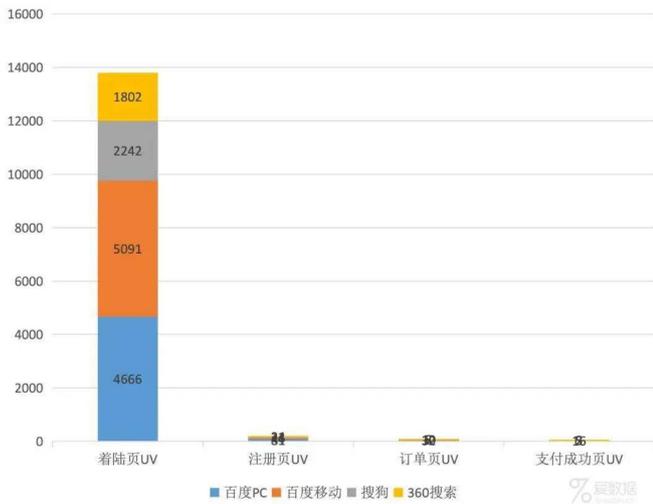
着陆页、注册页、订单页和支付成功页的总数据漏斗模型如下：



各渠道漏斗数据如下：



各渠道具体数据如下：



针对着陆页、注册页、订单页和支付成功页分别观察四个渠道两周内具体时间的走势图，如下：



## 五、提出假设并分析其原因

通过对以上的数据情况分析可以看出各渠道存在以下几点问题，对其中问题提出假设并分析其原因：

### 1. 整体渠道问题

(1)整体渠道来看，根据图1整体的漏斗图可以看出着陆页到注册页的流失异常严重，呈断崖式流失，其点击报名率只有1.25%。

#### 分析原因:

- 通过几个渠道的数据验证，发现几个渠道的着陆页到注册页的流量流失都十分严重，证明这不是个别情况，而是通病，可能是页面内容的吸引力不高。

- 通过对着陆页的观察，发现页面信息量较大较长，会有很多屏才可以看完，但是只有最后才有立即报名的按钮，跳转注册页，这样的话用户的浏览路径较长，在某一时刻产生报名冲动时没有入口。

#### 解决方案:

- 着陆页的页面信息、结构进行优化，增加吸引力。
- 进行A/B测试，在首屏增加立即报名的按钮，并且在用户浏览页面内容时保证报名入口常态悬浮屏幕底部，确保在任何时候用户可以进行点击转化。

(2)整体的注册转化率都不高，均值只有33.14%

#### 分析原因:

- 从用户从注册页面→手机账号/验证码→点击注册→订单确认的过程中可以看出流失过多，流失率达到将近6成。可能出现的原因在于注册流程过于复杂；
- 验证码接收时间过久或者接收成功率较低。

#### 解决方案:

- 简化注册步骤，确保只需要手机号和验证码就可以完成。
- 完善验证码接收成功率，和接收时间，确保多数人可以在短时间收到验证码。

#### 佐证数据:

- 进行用户访谈，观察其注册流程中产生的问题并进行记录；

- 对验证码的收发进行监测，计算其成功发送率。

(3)所有渠道的订单转化率都低，平均值不足50%，在最后一步流失同样严重；

#### 分析原因:

- 通过观察，在着陆页中没有明确价格标明，几百上千的价格对于用户来说心里决定成本较高；

- 支付流程过于复杂，操作难度较大；

- 支付方式单一。

#### 解决方案:

- 在着陆页首屏和报名按钮旁标明价格，让用户提前有心理预期；

- 优化支付流程，增加主流支付方式，比如：微信支付、支付宝等方式。

(4)所有渠道的页面跳出率都低，平均值达到57.54%，相当于每两个人中就有一人流失；

#### 分析原因:

- 着陆页内容吸引力不够，导致用户流失跳出；

- 页面内容过大，导致页面加载速度过慢，内容没有全部加载出来前用户就失去耐心关闭页面了；

- 用户匹配度不高，用户画像不精准。

#### 解决方案:

- 优化着陆页内容，突出卖点，增加吸引力；

- 前端页面进行质量压缩，保证正常网速下可以较快加载页面，全部页面加载时间进行优化；

- 优化搜索关键词，提高用户匹配度。

(5)所有渠道中百度（PC、移动）的用户停留时间和咨询占比均高于平均值，远远高于其他两个渠道。

#### 分析原因:

- 着陆页在搜狗和360搜索的布局出现问题，导致浏览体验差；

- 百度渠道的用户质量高于搜狗和360，匹配到的用户较为精准。

#### 解决方案:

- 对搜狗和360渠道用户进行调研，观察其页面操作路径，并进行对应优化

- 对这两个渠道的用户进行抽样访谈，看其对运营课程的感兴趣程度

## 2. 百度PC渠道分析

通过对几张UV走势图观察分析看出百度PC的着陆页、注册页、订单页、支付成功页的流量整体走势相同，波动较大。

### 分析原因:

- 在工作日用户上网习惯以PC为主，周末以手机为主导致。

### 解决方案:

· 运营可以对用户的使用习惯进行检测，精确到小时，观察用户是否是在工作时间流量较大，确认后可以有针对性的进行分时间段的广告投放，来优化投放效果。

## 3. 百度移动渠道分析

百度移动的点击报名率极低，低于均值，约是百度PC的50%；注册转化率也是四个渠道最低的，远远低于均值。

### 分析原因:

- 根据着陆页页面观察，内容较多较复杂，立即报名按钮在移动端也较小，不易点击，导致用户放弃流失；
- 注册流程出现问题或者过于复杂，用户长时间无法接收验证码；
- 百度移动端的点击报名率和注册转化率都是最低的，但是其订单转化率较高相对而言，可以猜测。

### 解决方案:

- 做移动端的着陆页和注册页的适配；
- 同样保证着陆页立即报名按钮常态悬浮屏幕底部；
- 优化注册流程，使注册流程简便；
- 测试验证码接收情况，保证其接收正常。

## 4. 搜狗渠道分析

搜狗渠道的点击报名率、订单转化率低于均值，用户平均停留时间最短，咨询占比也很低。

### 分析原因:

- 点击报名率低于均值除了有之前提到的原因外，猜测应该与搜狗本身渠道的流量较小有关，是百度PC的50%，百度移动的40%；
- 订单转化率较低猜测是由于搜狗浏览器支持的支付方式有关，并且其支付流程是否过于复杂；
- 从平均时长和较低的咨询占比可以看出搜狗的用户对其着陆页内容十分不感兴趣，用户群体不符合；
- 浏览器本身广告较多，右下角弹出后遮挡咨询浮窗。

### 解决方案:

· 对搜狗的用户进行用户调研，分析其用户画像，看是否与之相匹配，如果匹配程度不高可以考虑减少或者停止广

告投入；

- 对用户进行访谈观察，并且亲自体验其支付流程，完成其中的问题，简化支付流程或者丰富支付方式；
- 有针对性的调整咨询浮窗位置。

### 佐证数据:

- 搜狗用户画像；
- 全流程录屏记录。

## 5. 360搜索渠道分析

360搜索的注册转化率较低，低于平均值；订单转化率远远低于其他渠道；用户平均停留时间较短；跳出率是最高的；咨询占比最低

### 分析原因:

- 同以上相同问题原因
- 从各项指标都低可以看出360搜索的用户群极不匹配，不是其受众。

### 解决方案:

- 优化支付流程、注册流程
- 对360搜索用户进行用户调研，看其用户画像的匹配程度，如重合度极低的话可以考虑放弃投放。

接下来的工作中，建议营销端重点关注哪个渠道？做什么动作？为什么？

1. 对于百度PC渠道可以增加广告位的投放和有时间段的针对性投放

理由：百度PC流量大，且各项指标的转化率均高于均值，如果再优化投放效果的话，可以使ROI最大化

2. 首先着重优化百度系渠道的着陆页结构和内容，提升点击报名率，这点是重中之重；其次简化注册流程，提升注册转化率。

理由：

· 因为从流量上来看，这两个渠道的流量最大，尤其是百度移动。

· 从页面停留时长和咨询占比看，这两个渠道的用户对页面内容更为感兴趣，利于转化。

3. 对于搜狗和360搜索这两个渠道可以考虑减少投放量或者不进行投放。

理由：经过上述分析，可以看出这两个渠道本身的流量较少，并且用户精准度不高，不是主要用户群，ROI比较低。

## / 20种数据可视化工具测评, 有一半你都不知道 /

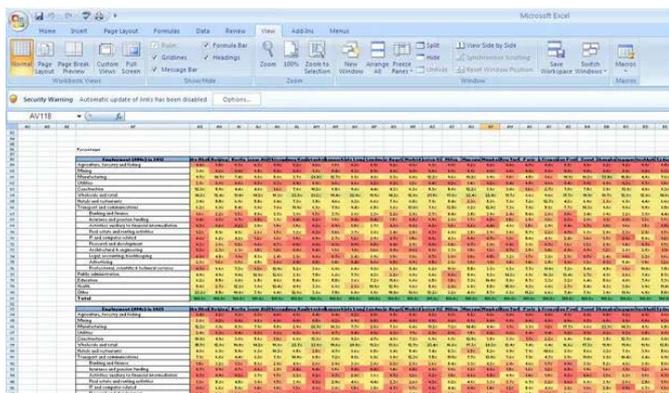
来源 / PLoB 编辑 / 协会会员处 李苗苗 日期 / 2020-09



如今学习应用数据可视化的渠道有很多, 你可以跟踪一些专家博客, 但更重要的一点是实践/实操, 你必须对目前可用的数据可视化工具有个大致了解。以下是Netmagazine列举的二十大数据可视化工具, 无论你是准备制作简单的图表还是复杂的图谱或者信息图, 这些工具都能满足你的需要。更加美妙的是, 这些工具大多免费。

## 一、入门级工具

## 1.Excel



Excel的图形化功能并不强大, 但Excel是分析数据的理想工具, 上图是Excel生成的热力地图。

作为一个入门级工具, Excel是快速分析数据的理想工具, 也能创建供内部使用的数据图, 但是Excel在颜色、线条和样式上可选择的范围有限, 这也意味着用Excel很难制作出能符合专业出版物和网站需要的数据图。但是作为一个高效的内部沟通工具, Excel应当是你百宝箱中必备的工具之一。

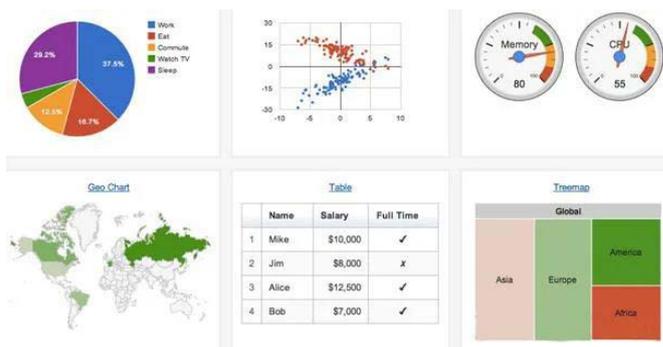
## 2.CSV/JSON

CSV(逗号分隔值)和JSON (JavaScript对象注释) 虽然不是真正的可视化工具, 但却是常见的数据格式。你必须理解

他们的结构, 并懂得如何从这些文件中导入或者导出数据。以下将要介绍的所有数据可视化工具都支持CSV、JSON中至少一种格式。

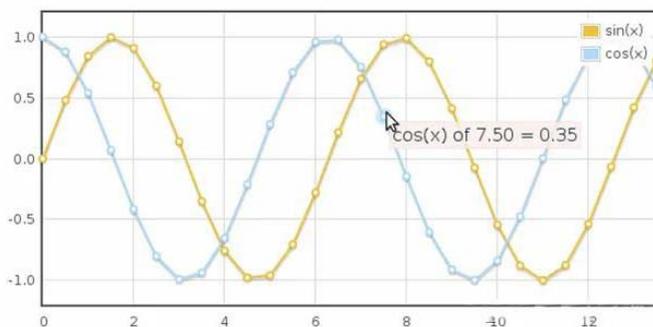
## 二、在线数据可视化工具

## 1.Google Chart API



Google Chart API工具集中取消了静态图片功能, 目前只提供动态图表工具。能够在所有支持SVG\Canvas和VML的浏览器中使用, 但是Google Chart的一个大问题是: 图表在客户端生成, 这意味着那些不支持JavaScript的设备将无法使用, 此外也无法离线使用或者将结果另存其他格式, 之前的静态图片就不存在这个问题。尽管存在上述问题, 不可否认的是Google Chart API的功能异常丰富, 如果没有特别的定制化需要, 或者对Google视觉风格的抵触, 那么你可以从Google Chart开始。

## 2.Flot



Flot是一个优秀的线框图图库, 支持所有支持canvas的浏览器 (目前主流的浏览器如火狐、IE、Chrome等都

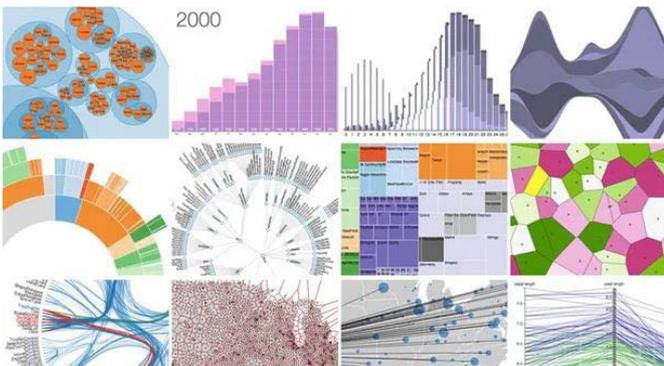
支持)。

### 3.Raphaël



Raphaël是创建图表和图形的JavaScript库，与其他库最大的不同是输出格式仅限SVG和VML。SVG是矢量格式，在任何分辨率下的显示效果都很好。

### 4.D3



D3 (Data Driven Documents) 是支持SVG渲染的另一种JavaScript库。但是D3能够提供大量线性图和条形图之外的复杂图表样式，例如Voronoi图、树形图、圆形集群和单词云等。虽然D3能够提供非常花哨的互动图表，但你在选择数据可视化工具时，需要牢记的一点是：知道在何时保持简洁。

### 5.Visual.ly

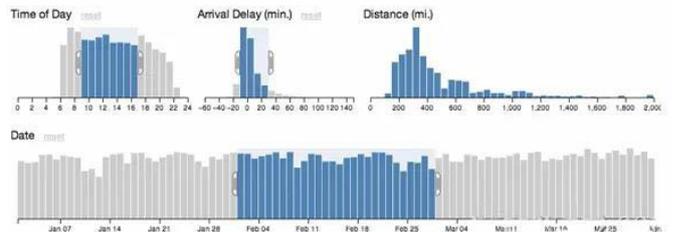


如果你需要制作信息图而不仅仅是数据可视化，目前也有大把的工具可用。Visual.ly就是最流行的一个选择。虽然Visual.ly的主要定位是：“信息图设计师的在线集市”，但是也提供了大量信息图模板。虽然功能还有很多限制，但是Visual.ly绝对是个能激发你灵感的地方。

## 三、互动图形用户界面 (GUI) 控制

如果数据可视化的互动性强大到可以作为GUI界面会怎样？随着在线数据可视化的发展，按钮、下拉列表和滑块都在进化成更加复杂的界面元素，例如能够调整数据范围的互动图形元素，推拉这些图形元素时输入参数和输出结果数据会同步改变，在这种情况下，图形控制和内容已经合为一体。以下这些工具能够帮你实现这些功能：

### 1.Crossfilter



当我们为方便客户浏览数据开发出更加复杂的工具时，我们已经能够创建出既是图表，又是互动图形用户界面的小程序。JavaScript库Crossfilter就是这样的工具。

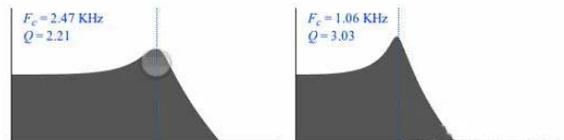
Crossfilter应用：当你调整一个图表中的输入范围时，其他关联图表的数据也会随之改变。

### 2.Tangle

The coefficients and transfer function are:

$$k_f = 0.35 \quad k_v = 0.453$$
$$H(z) = \frac{0.122}{1 - 1.719z^{-1} + 0.842z^{-2}}$$

Some example frequency responses:



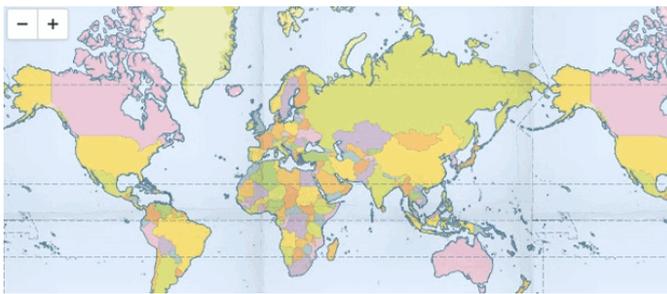
JavaScript库Tangle进一步模糊了内容与控制之间的界限。在下图这个应用实例中，Tangle生成了一个负载的互动方程，读者可以调整输入值获得相应数据。

## 四、地图工具

地图生成是web上最困难的任务之一。Google Maps的出现完全颠覆了过去人们对在线地图功能的认识。而Google发布的Maps API则让所有的开发者都能在自己的网站中植入地图功能。

近年来，在线地图的市场成熟了很多，如果你需要在数据可视化项目中植入定制化的地图方案，目前市场上已经有很多选择，但是知道在何时选择何种地图方案则成了一个很关键的决策。地图方案看上去功能都很强大，但是切忌：“有了一把锤子，看什么都像钉子。”

### 1.Modest Maps



顾名思义，Modest Maps是一个很小的地图库，只有10KB大小，是目前最小的可用地图库。这似乎意味着Modest Maps只提供基本的地图功能，但是不要被这一点迷惑了。在一些扩展库的配合下，例如Wax，Modest Maps立刻会变成一个强大的地图工具。

### 2.Leaflet



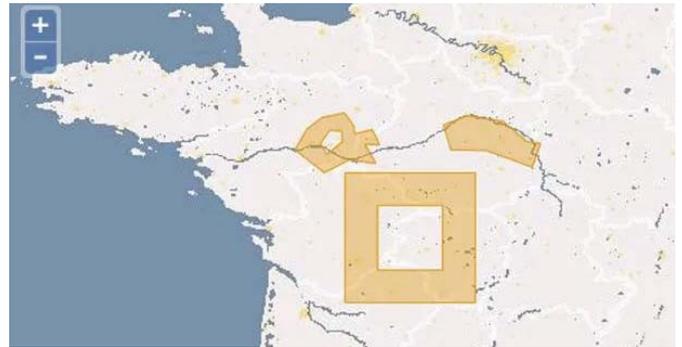
CloudMade团队为大家带来了Leaflet，这是另外一个小型化的地图框架，通过小型化和轻量化来满足移动网页的需要。Leaflet和Modest Maps都是开源项目，有强大的社区支持，是在网站中整合地图应用的理想选择。

### 3.PolyMaps



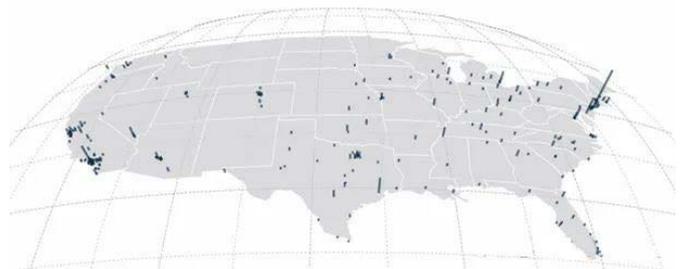
Polymaps是另外一个地图库，但主要面向数据可视化用户。Polymaps在地图风格化方面有独到之处，类似CSS样式表的选择器，是不可错过的好东西。

### 4.OpenLayers



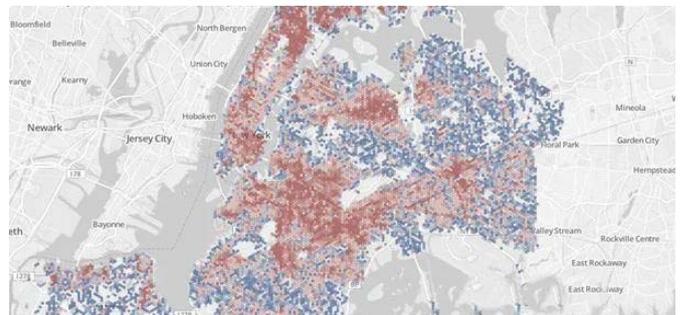
OpenLayers可能是所有地图库中可靠性最高的一个。虽然文档注释并不完善，且学习曲线非常陡峭，但是对于一些特定的任务来说，OpenLayers无可匹敌。例如能够提供一些其他地图库都没有的特殊工具。

### 5.Kartograph



Kartograph的标记线是对地图绘制的重新思考，我们都已经习惯了莫卡托投影，但是Kartograph为我们带来了更多的选择。如果你不需要调用全球数据，而仅仅是生成某一区域的地图，那么Kartograph将使你脱颖而出。

### 6.Cartodb



CartoDB是一个不可错过的网站。你可以用CartoDB很容易就把表格数据和地图关联起来，这方面CartoDB是最优秀的选择。例如，你可以输入CSV通讯地址文件，CartoDB能将地址字符串自动转化成经度/维度数据并在地图上标记出



来。目前CartoDB支持免费生成五张地图数据表，更多使用需要支付月费。



(随着iPad3等高清移动设备的普及) web开发的一个最新趋势是将符号字体与字体整合(把符号变成字体)，创建出漂亮的矢量化图标。在这些新型字体中，例如FF Chartwell和Chartjunk是专门用来显示图表和图形的。他们与OpenType碰到的问题一样，就是不能被所有的浏览器支持，但是不久的将来这些矢量字体将是数据可视化工作中需要考到的因素。

## 五、进阶工具

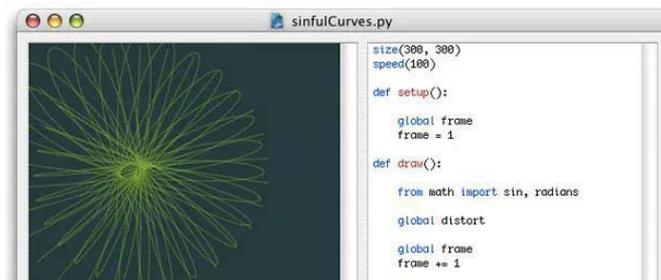
如果你准备用数据可视化做一些“严肃”的工作，那么你可能不会对在线可视化工具或者web小程序有太大兴趣，你需要的是桌面应用和编程环境。

### 1.Processing



Processing是数据可视化的招牌工具。你只需要编写一些简单的代码，然后编译成Java。目前还有一个Processing.js项目，可以让网站在没有Java Applets的情况下更容易地使用Processing。由于端口支持Objective-C，你也可以在iOS上使用Processing。虽然Processing是一个桌面应用，但也可以在几乎所有平台上运行，此外经过数年发展，Processing社区目前已近拥有大量实例和代码。

### 2.NodeBox



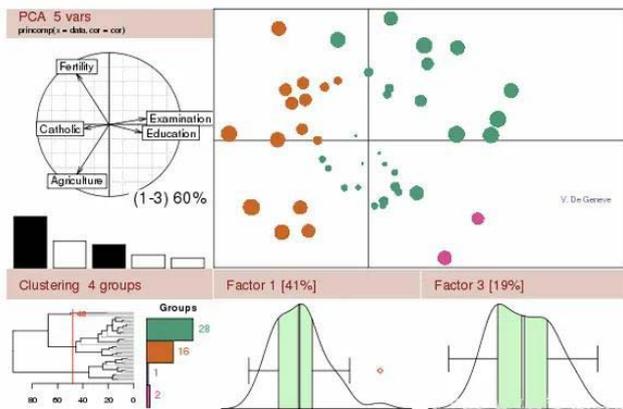
NodeBox是OS X上创建二维图形和可视化的应用程序。你需要了解Python程序，NodeBox与Processing类似，但是没有Processing的互动功能。

## 六、专家级工具

与Excel相对的是专业数据分析工具。如果你是一个专业的数据分析师，那么你就必须对下面将要介绍的工具有所了解(如果不是精通的话)。

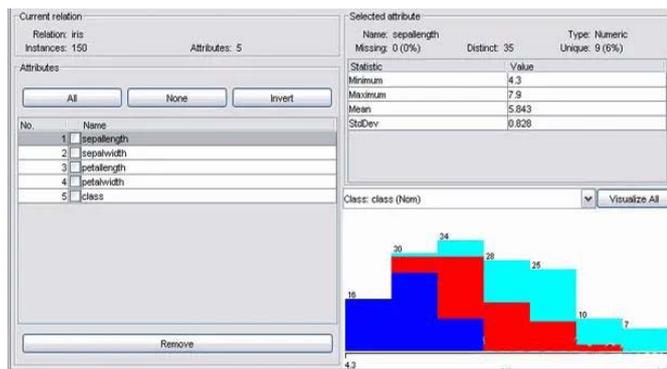
众所周知，SPSS和SAS是数据分析行业的标准工具，但是这些工具的费用不菲，只有大型组织和学术机构才有机会使用，下面我们介绍几种免费的替代工具，这些开源工具的共同特征是都有强大的社区支持。开源分析工具性能不输老牌专业工具，插件的支持甚至更好。

## 1. R



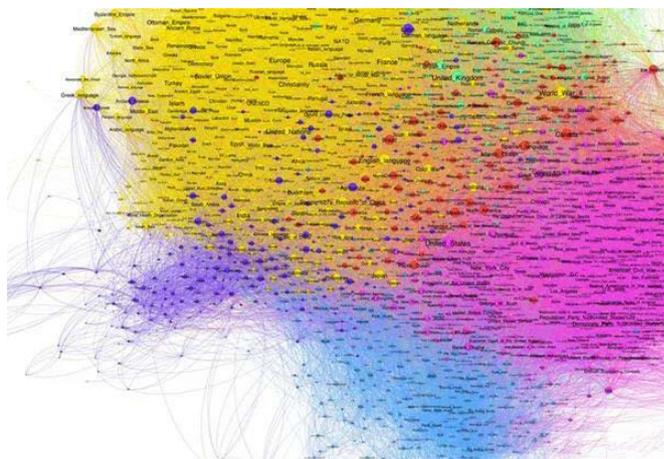
作为用来分析大数据集的统计组件包，R是一个非常复杂的工具，需要较长的学习实践，学习曲线也是本文所介绍工具中最陡峭的。但是R拥有强大的社区和组件库，而且还在不断成长。当你能驾驭R的时候，一切付出都是物有所值的。

## 2. Weka



当你成长成为一名数据科学家的时候，你需要将个人能力从数据可视化扩展到数据挖掘领域。Weka是一个能根据属性分类和集群大量数据的优秀工具，Weka不但是数据分析的强大工具，还能生成一些简单的图表。

## 3. Gephi



Gephi是进行社交图谱数据可视化分析的工具，不但能处理大规模数据集并生成漂亮的可视化图形，还能对数据进行清洗和分类。Gephi是一种非常特殊的软件，也非常复杂，先于他人掌握Gephi将使你一骑绝尘。

## / Excel做"回归分析"，你可能都没玩儿过！ /

来源 / 数据说 编辑 / 协会会员处 李苗苗 日期 / 2020-09

网站分析中专业的工具除了Google Analytics, Adobe Sitecatalyst, Webtrends, 腾讯分析和百度统计等外，我想最常用的数据处理工具就是Excel了，Excel里头最基础的就是运算和图表的制作，稍微高级一点就是函数和数据透视表的使用了，当然你可能还会想到VBA和宏，但估计很少高手会使用这些高级的功能。

那对于高级的数据分析而言，也就是涉及统计学的专业

分析方法和原理的时候，是不是就一定得求助于SPSS,SAS这类专业的分析工具呢？数据分析从低级到高级层次的跳跃过程中有没有可以起承接作用的工具呢？

其实是有的，这就是Excel的数据分析功能。高级的数据分析会涉及回归分析、方差分析和T检验等方法，不要看这些内容貌似跟日常工作毫无关系，其实往高处走，MBA的课程也是包含这些内容的，所以早学晚学都得学，干脆就提前了解



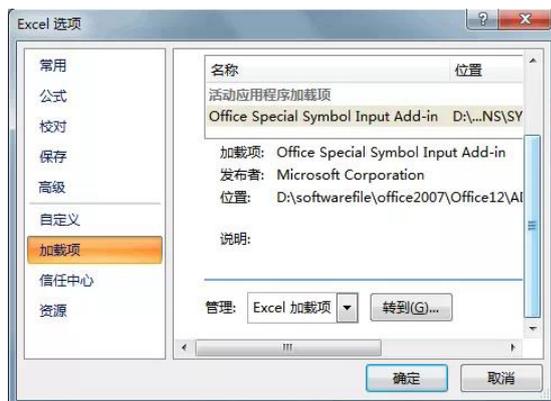
吧，请查看以下内容。

在使用之前，首先得安装Excel的数据分析功能，默认情况下，Excel是没有安装这个扩展功能的，安装如下所示：

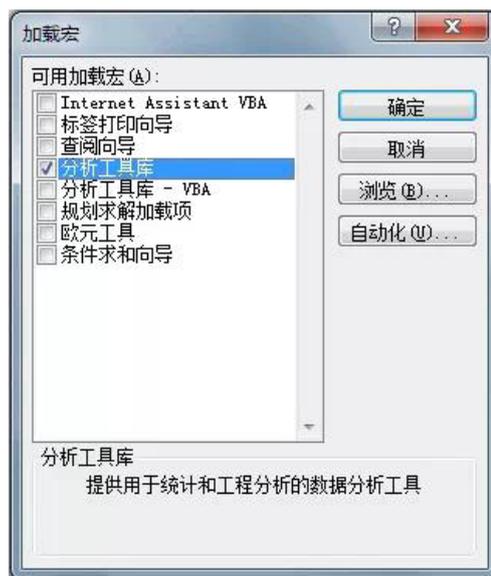
1. 鼠标悬浮在Office按钮上，然后单击【Excel选项】



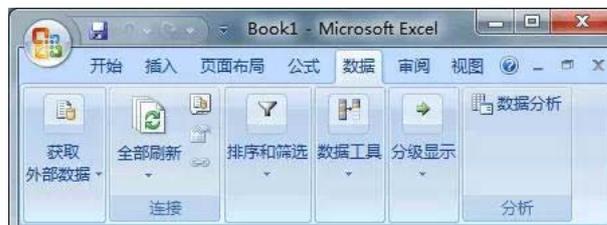
2. 找到【加载项】，在管理板块选择【Excel加载项】，然后单击【转到】



3. 选择【分析工具库】，单击【确定】



4. 安装完后，就可以【数据】板块看到【数据分析】功能，如下所示



安装完后，首先来了解一下回归分析的内容。

### 回归分析

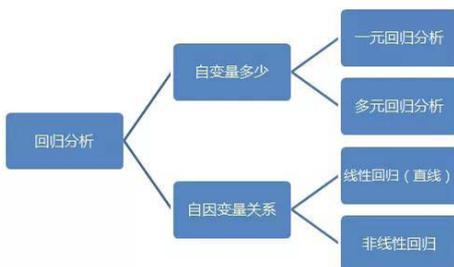
在详细进行回归分析之前，首先要理解什么叫回归？实

实际上，回归这种现象最早由英国生物统计学家高尔顿在研究父母亲和子女的遗传特性时所发现的一种有趣的现象：

身高这种遗传特性表现出”高个子父母，其后代身高也高于平均身高；但不见得比其父母更高，到一定程度后会往平均身高方向发生’回归’”。这种效应被称为”趋中回归”。现在的回归分析则多半指源于高尔顿工作的那一整套建立变量间的数量关系模型的方法和程序。这里的自变量是父母的身高，因变量是子女的身高。

百度百科对于回归分析的定义是: 回归分析 (regression analysis)是确定两种或两种以上变数间相互依赖的定量关系的一种统计分析方法。运用十分广泛：

- 回归分析按照涉及的自变量的多少，可分为一元回归分析和多元回归分析。
- 按照自变量和因变量之间的关系类型，可分为线性回归分析和非线性回归分析。



### 应用案例

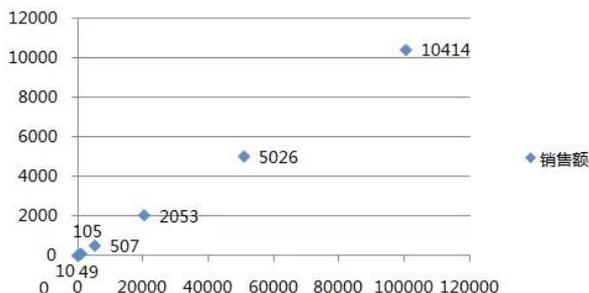
这里举个电商的例子：电子商务的转换率是一定的，网站访问数一般正比对应于销售收入，现在要建立不同访问数情况下对应销售的标准曲线，用来预测搞活动时的销售收入，如下所示：

访问数	销售额
0	0
100	10
489	49
1010	105
5230	507
20400	2053
51000	5026
100602	10414

1. 首先，利用散点图描绘图形



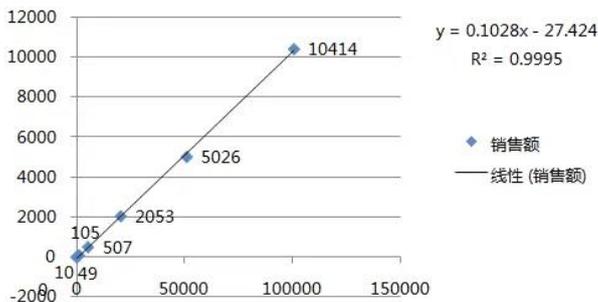
### 销售额



2. 添加趋势线，并且显示回归分析的公式和R平方值



### 销售额

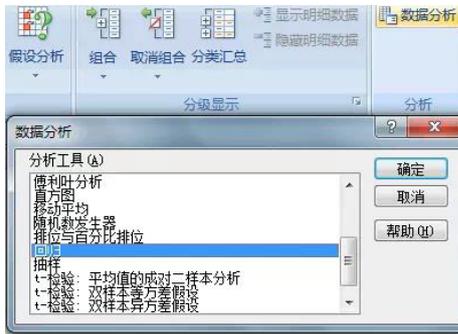


从图得知，R平方值=0.9995，趋势线趋同于一条直线，公式是： $y=0.01028x-27.424$

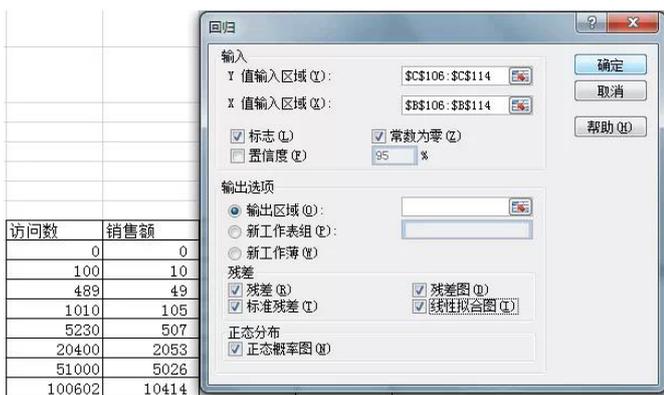
R平方值是介于0和1之间的数字，当趋势线的R平方值为1或者接近1时，趋势线最可靠。因为 $R^2 > 0.99$ ，所以这是一个线性特征非常明显的数值，说明拟合直线能够以大于99.99%地解释、涵盖了实际数据，具有很好的一般性，能够起到很好的预测作用。

3. 使用Excel的数据分析功能

(1) 点击【数据分析】，在弹出的选择框中选择【回归】，然后点击【确定】



(2) 【X值输入区域】选择访问数的单元格，【Y值输入区域】选择销售额的单元格，同时勾选如下所示的选项，包括残差、标准残差、残差图、线性拟合图和正态概率图。

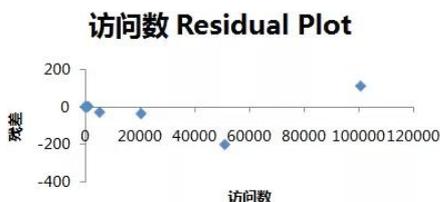


访问数	销售额
0	0
100	10
489	49
1010	105
5230	507
20400	2053
51000	5026
100602	10414

(3) 以下内容是残差和标准残差

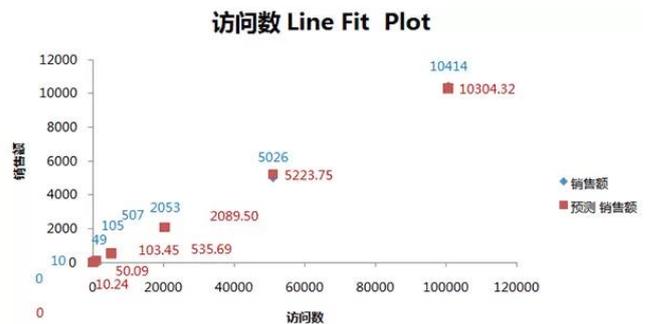
SUMMARY OUTPUT						
回归统计						
Multiple R		0.99980694				
R Square		0.999613917				
Adjusted R Square		0.856756775				
标准误差		87.30185573				
观测值		8				
方差分析						
	df	SS	MS	F	Significance F	
回归分析	1	138132873	138132873	18123.83477	1.13286E-11	
残差	7	53351.2981	7621.614014			
总计	8	138186224.3				
RESIDUAL OUTPUT		PROBABILITY OUTPUT				
观测值	预测 销售额	残差	标准残差	百分比排位	销售额	
1	0	0	0	6.25	0	
2	10.24	-0.2426567	-0.0029714	18.75	10	
3	50.09	-0.9532579	-0.011673	31.25	49.1333333	
4	103.45	1.88250078	0.02305195	43.75	105.3333333	
5	535.69	-28.357611	-0.3472499	56.25	507.3333333	
6	2089.50	-36.168631	-0.4428989	68.75	2053.333333	
7	5223.75	-198.22158	-2.4272998	81.25	5025.533333	
8	10304.32	109.282518	1.33820668	93.75	10413.6	

(4) 以下是残差图



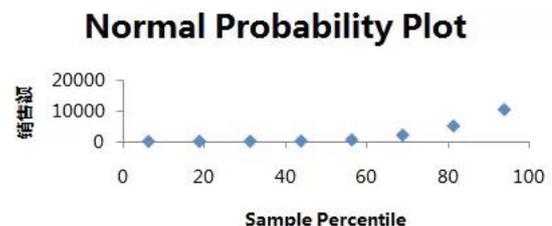
残差图是有关于实际值与预测值之间差距的图表，如果残差图中的散点在中轴上下两侧分布，那么拟合直线就是合理的，说明预测有时多些，有时少些，总体来说是符合趋势的，但如果都在上侧或者下侧就不行了，这样有倾向性，需要重新处理。

(5) 以下是线性拟合图



在线性拟合图中可以看到，除了实际的数据点，还有经过拟和处理的预测数据点，这些参数在以上的表格中也有显示。

(6) 以下是正态概率图



正态概率图一般用于检查一组数据是否服从正态分布，是实际数值和正态分布数据之间的函数关系散点图，如果这组数值服从正态分布，正态概率图将是一条直线。回归分析不一定得符合正态分布，这里只是仅仅把它描绘出来而已。

以上数据表格和图表都说明公式 $y=0.01028x-27.424$ 是一个值得信赖的预测曲线，假设搞活动时流量有50万访问数的话，那么预测销售将是51373，如下图所示：

访问数	销售额		
0	0		
100	10		
489	49		
1010	105		
5230	507		
20400	2053		
51000	5026		
100602	10414		
500000	51373	公式: $y=0.01028x-27.424$	

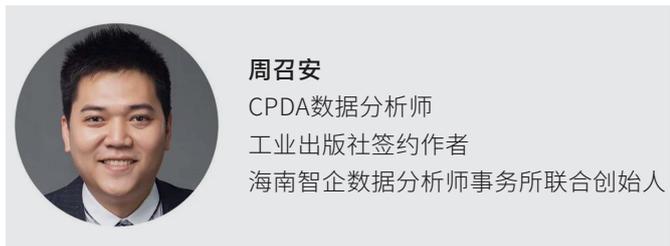
## / CPDA学员专访

# 周召安：成就源于平凡的选择和不放弃的努力 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-09

“成就源于平凡的选择和不放弃的努力。”

这句话是《给新手的企业数据分析实战》一书作者周召安的座右铭，也是他本人这些年来对于CPDA数据分析师学习和从业经历的感悟和总结。



**周召安**  
CPDA数据分析师  
工业出版社签约作者  
海南智企数据分析师事务所联合创始人

周召安带领他的数据分析师团队，常年致力于为企业（委托方）提供数据处理、数据挖掘、数据可视化等数据分析技术服务，多年间解决了多起企业数据运用的难题，大大增强了企业的人力效能和经营效能。他本人也获得了由中国商业联合会数据分析专业委员会颁发的“卓越数据分析师”称号。

2011年毕业的他，经过长达1年的考量后于2014年通过参加CPDA的学习培训，考取了CPDA数据分析师，经猎头推荐进入深圳国内某知名集团任职数据分析师。

现在海南当地与同为CPDA数据分析师的合伙人共同创办了“海南智企数据分析师事务所”，常年为服装、医美、餐饮、快消品等多行业企业提供数据分析服务。



图：周召安代表海南智企数据分析师事务所做行业经验分享



图：周召安团队为合作学校老师做数据分析培训

2018年周召安受电子工业出版社邀约，将其多年的数据分析从业经验撰写成企业级数据分析丛书，共分两部。历时两年，第一部《给新手的企业数据分析实战》已于今年6月完成，并于9月正式在京东、天猫等电商平台上架销售。



这位2014级的CPDA学长，如今已在数据分析行业小有成就。国庆假期之前，数据君对师兄进行了远程访谈，聊了聊当年他参加CPDA学习的心路历程以及这些年在行业中积累的经验心得。

### “圈外人”的触电

谈到当初为什么选择参加CPDA的培训学习这一话题时，周召安笑说自己其实是个“圈外人”，与CPDA的相识真的是奇妙的缘分：“我本人并非是学数据分析相关专业出身，而是学的计算机科学与技术，主攻数据库。”

周召安大学毕业时并没有很幸运找到和自己所学专业对

口的工作，而是先去做了电商运营，然而做电商运营需要数据分析支持，他和数据分析的缘分就此开始了。由于当时任职的公司不大，没有设立专门的数据分析岗位，所以数据分析这个工作只能由自己这个运营人员来做，这让周召安萌生了想系统去学习的想法。但是在他的认知里，系统学习一门技术只能通过参加学校的学习，但是学校教学的内容在系统性上显得比较散，且偏技术语言工具，外行人是听不懂的。



图：大学毕业刚刚参加工作的周召安

找到适合自己的培训课程，成为了当时周召安决心学习数据分析要解决的第一要务。有了这个念头，周召安就开始上网搜索和数据分析相关的培训课，因为本身从事电商运营，对于“关键词”的选择有着一套自己的经验，很快就找到CPDA数据分析师培训。

周召安在回忆这段往事时说：“其实当时是有对比过市场上很多跟数据分析相关的培训课程，但看到CPDA课程的时候，一下就被吸引了。于是我用了较长的一段时间对咱们的课程体系做了深入详细的了解，最终确定了报名CPDA。”

### 最重要的是系统化学习

CPDA课程带给周召安最大的收获，就是让他对数据分析有了一个系统化的认知。作为一个没有数据分析相关工作经验的外行，周召安明白自己要想进入这个全新的领域，必须要了解全面系统的知识体系，得先知道数据分析这头“大象”大体的框架是怎样的。



图：CPDA广州8期班全体学员合影

在没有学习前，他认为数据分析就是简单将他在运营工作中所涉及的数据进行统计分析。但是通过系统的学习后，他对数据分析有了新的理解：想要成为一名合格的认证数据分析师，需要立体的分析，全面的考量。这也就是CPDA一直倡导的大数据思维，构建起全新的思维模式就可以为我们的工作决策提供强大的分析支撑，而作为数据分析师及从业者，自身能力的培养提升，也应当围绕着这方面进行积累。

### 加入CPDA大家庭的收获

“我家是海南的，当时海南并没有授权中心，所以只能到广州去学习。那时刚毕业没多久，基本都是月光族，只能跟家里面要钱，然后报名，脱产去学习。学习期间，我就住在朋友那边，待了整整一个月。”周召安回忆着2014年参加CPDA学习时的情景。

参加CPDA课程学习后，虽然当时通过了考试，也拿到了《CPDA数据分析师》证书，但因对数据分析的继续教育理解不够深刻，再加上当时整体环境还不像现在这样成熟，2015年的周召安曾一度陷入了迷茫，不知道自己作为数据分析师接下来要往哪里发展，于是那一年他选择留在了广州。



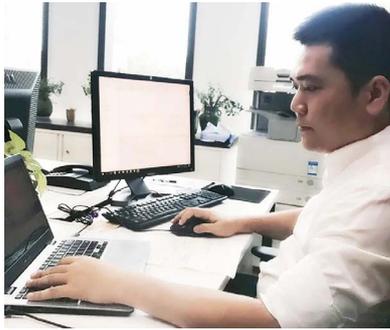
图：CPDA学员聚餐联谊

一个偶然的的机会，他遇到了同为CPDA数据分析师的姐姐，通过深度沟通后他有了新的方向：探寻数据分析工具的应用。他开始到处寻味认识CPDA数据分析师。之后他又认识了来自河南的分析师科研，针对数据分析的实际应用价值与同行进行了更深入的探讨。再到后来，又结识了来自深圳、成都各地的CPDA数据分析师，并以同样的方式，获取了新的技能。周召安表示：“通过CPDA这个大家庭，他结识了很多志同道合的小伙伴，因为有了大家彼此支持和鼓励，才走的更远更好！”

### 新的起点：回乡创业

2017年，也就是CPDA结业的第二年，周召安做了一个决

定，为了更好的成家，从深圳离开，回到自己的家乡海南，回到海南没有找到数据分析师岗位工作，于是选择创业：与结识的CPDA数据分析师一起合伙成立了事务所：海南智企数据分析事务所，为企业提供数据分析相关服务。



图：数据分析师的工作日常

创业并没有想象的那么容易，周召安这次人生的转型也经历了波折坎坷。由于缺少宣传沟通渠道、聘任方提出的需求与自身能力不匹配，以及事务所自身业务与经营模式不匹配等诸多因素，事务所成立前8个月都没有业务。即使在这样艰难的时期，他仍然选择不放弃，而是带领团队不断试错、总结，同时制定了止损时间，时间没到，坚决不放弃。



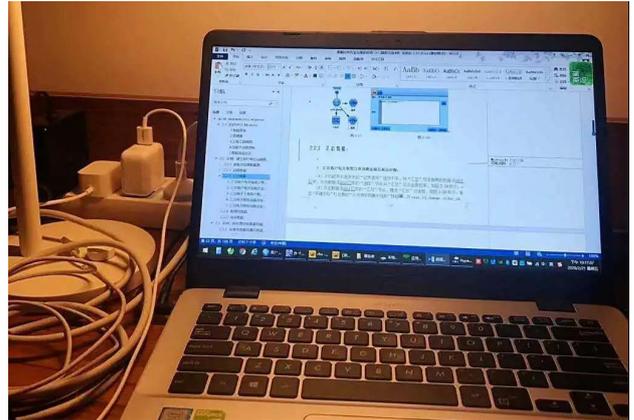
图：全员头脑风暴

这样的境况直到2018年末，才迎来了转机。事务所慢慢被诸多企业认知、认可，陆续接到了一些常年事务的委托。现在海南智企数据分析师事务所已经发展成为一家集培训、服务、应用为一体，提供企业级数据分析咨询服务的事务所。周召安说自己很感谢那段经历，也感谢自己的坚持，没有轻易放弃。

**《给新手的企业数据分析实战》  
写给像他曾经一样的“圈外人”**

这本书是周召安多年来从事数据分析行业的经验总结，书中尽可能的还原了他与本书另一位作者从CPDA认证培训结业后所遇到的人和事。本书通过技术实现——方法论应用——

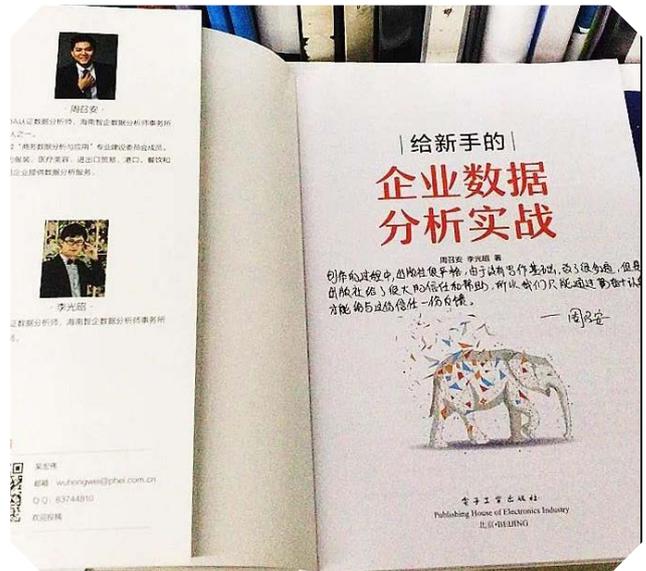
思路设计的主干线，以及场景对话的方式，简单明了的阐述了处理企业数据分析不可或缺的基础知识和应当专注的方向。周召安说这个逻辑和思考方式，是当年参加CPDA培训学到的。



图：《给新手的企业数据分析实战》的创作环境

谈到什么样的读者适合看这本书，会不会感觉专业性很强？周召安给出答案，一是没有接触过数据分析的人；二是没用过专业数据分析工具的人。而这两类人身上又恰恰有他曾经的样子：数据分析师新人。

这本书历时两年完成，汇集了这些年周召安在这个行业的实战从业经验以及当年他作为一个“圈外人”选择参加CPDA学习到现在一路走来的心路历程。



图：签名版《给新手的企业数据分析实战》

周召安在和数据君诉说自己从开始学习CPDA、到之后

的创业、再到现在出书，这一路走来是真的不容易。如今的他在数据分析行业早已不再是小白，能够快速成功入行，很感谢当年在CPDA课堂学习的第一课：数据分析这头“大象”原来是长这样的。而对于他今天的成就，用他自己的话说：

成就源于平凡的选择和不放弃的努力。衷心祝愿师兄越走越远，祝海南智企数据分析师事务所蒸蒸日上！

## / 山东智谷数据分析师事务所 /

来源 / 山东智谷数据分析师事务所 编辑 / 协会会员处 李苗苗 日期 / 2020-09

山东智谷数据分析师事务所，是山东省知名数据分析师事务所。事务所由山东众志电子有限公司发起成立。

目前服务于省内外政府、事业单位和大中型企业，为各行业领域提供大数据分析体系建设服务、大数据环境搭建、区块链应用开发、数据汇集整理、主题数据库建设、大数据分析建模、实时流数据处理、舆情分析等服务。

### 一、事务所资质

事务所是中国商业联合会数据分析专业委员会（简称“中数委”）的事务所会员单位，具备中数委颁发的会员执业资质证书。

2019年5月10日，中数委通过事务所备案及入会申请，官网公示批复；

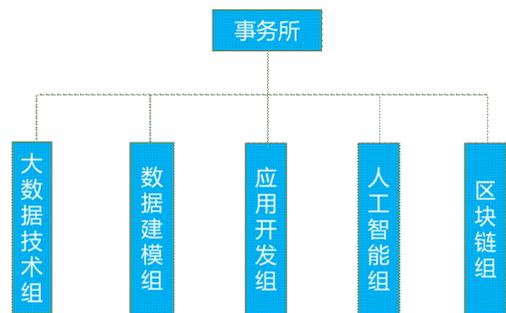
2019年5月17日，中数委颁发中国数据分析行业“会员执业资质证书”。



### 二、团队力量

目前，事务所拥有专业的CPDA数据分析师，具有丰富计算机科学和统计学数据分析经验；事务所还拥有专业从事大数据分析挖掘、模型算法研究和平台软件研发的复合型人才。

- 工信部和CPDA数据分析师；
- 大数据建模与分析挖掘高级工程师；
- 大数据应用开发高级工程师。



组织架构

### 三、技术构成

事务所具备使用Python、R语言进行数据分析的能力，可以搭建基于开源Hadoop生态圈组件的大数据分布式存储和计算集群，使用基于内存计算的Spark组件进行大数据应用开发和数据分析建模；为企业事业单位提供数据分析挖掘、数据分析报告、区块链应用开发服务和定制化企业上云服务。



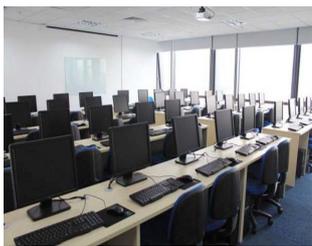
### 四、技术力量

事务所依托山东众志电子有限公司承建运营的泰安市云计算中心，拥有优质的存储计算资源、多领域的行业云平台和专业的运维服务，业务遍及山东省农业科学院奶牛研究中心、泰安市大数据局、泰安市公安局泰山区分局、泰安市人防办等部门。



### 五、部分项目案例

#### (一) 驻场数据服务 (大数据局、人防办)



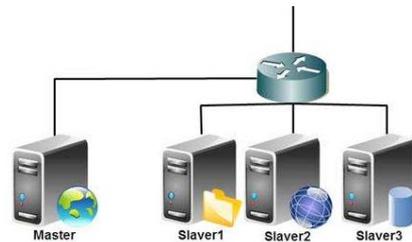
#### (二) 大数据分析报告 (高校教务处与信工学院)



#### (三) 奶牛育种大数据平台 (农科院)



#### (四) 大数据集群部署和数据分析 (审计局)



#### (五) 企业舆情分析 (公安局)



#### (六) 亩产效益分析系统 (工信局)



#### (七) 智慧社区警务云平台 (公安局)



### 六、部分模型应用案例

#### (一) RFM客户价值分析

RFM模型的最终目的是为了在经营过程中区别出客户类群，并在此基础上针对不同类型客户进行精准化营销，进而提



升客户价值，实现企业利益最大化。

R (近度)	F (频度)	M (额度)	客户类型
高	高	高	高价值客户
低	高	高	重点保持客户
高	低	高	重点发展客户
低	低	高	重点挽留客户
高	高	低	一般价值客户
低	高	低	一般保持客户
高	低	低	一般发展客户
低	低	低	潜在客户

(二) 文本挖掘应用于情感分类



使用朴素贝叶斯算法为点评文本做情感分类。

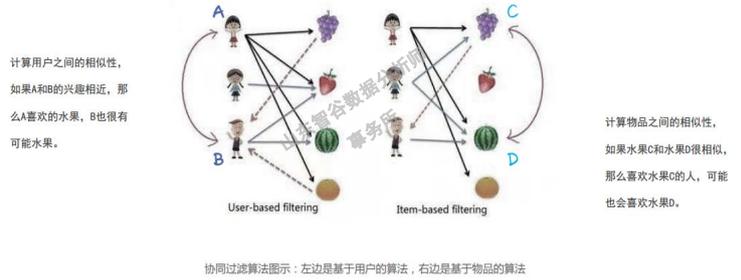
	comment	star	sentiment	snlp_result	cut_comment	nb_result
0	口味：不知道是我口高了，还是这家真不怎么样。??我感觉口味确实很一般很一般。上菜相当快，我敢...	2	0	0	口味：不知道是我口高了，还是这家真不怎么样。??我感觉口味确实很一般很一般。上菜相当快，我敢...	1
1	菜品丰富质量好，服务也不错！很喜欢！	4	1	1	菜品丰富质量好，服务也不错！很喜欢！	1
2	说真的，不晓得有人排队的理由，香精香精香精香精，拜拜！	2	0	1	说真的，不晓得有人排队的理由，香精香精香精香精，拜拜！	0
3	菜量实惠，上菜还算比较快，疙瘩汤喝出了秋日的暖意，烧茄子吃出了大板烧的味道，想吃土豆片也是口...	5	1	0	菜量实惠，上菜还算比较快，疙瘩汤喝出了秋日的暖意，烧茄子吃出了大板烧的味道，想吃土豆片也是口...	1
4	先说我是哪家风荷园开业就一直在这里吃??每次出去回来总想吃一回??有时觉得外面的西式晚餐...	4	1	0	先说我是哪家风荷园开业就一直在这里吃??每次出去回来总想吃一回??有时觉得外面的西式晚餐...	0

(三) 人员画像

对于预警嫌疑人对象，显示其个人基本信息，字段包括姓名、年龄、性别、工作、学历、住址、是否有前科等信息。

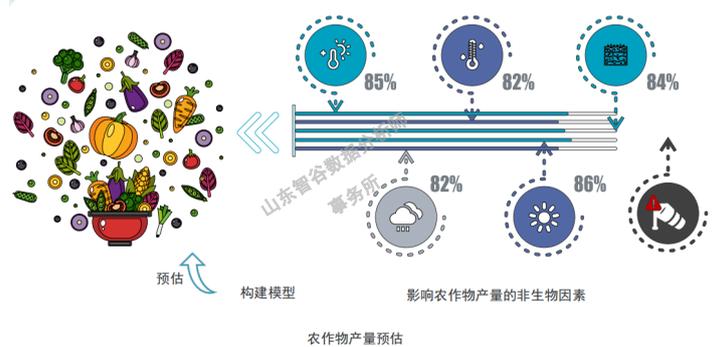


(四) 产品推送



协同过滤算法图示：左边是基于用户的算法，右边是基于物品的算法

(五) 产量预估



业务联系

联系人：苏航  
联系电话：0538-8932988 / 188 5383 9679  
办公地址：山东省泰安高新区南天门大街1110号



CPDA<sup>®</sup> 数据分析师  
Certified Projects Data Analyst

# 不甘平凡

UNWILLING  
ORDINARY



关注CPDA数据说



ON MY WAY