



AsiaAnalytics



**Data. Insights. Performance.**



## 数据分析现状

2013 年是大数据元年，数据就是资源，数据就是竞争优势，数据就是财富，对数据的分析和处理已刻不容缓，特别是社会化媒体的出现和普及，更是引起了数据爆炸性地增长；各个公司都希望从已有的规则数据和非规则数据中能挖掘出有价值的信息，并预测客户的行为需求和走向，以便最优化公司资源配置，从而达到公司利益的最大化。

**面对这场无硝烟的数据战，您是否做好了充足的准备？请您先思考以下几个问题：**

- 您了解并试图动用全公司的资源来解决公司现存的问题吗？
- 公司的数据量是否充足到可以被用来做数据分析和挖掘？
- 您对数据分析、数据挖掘的方法论和工具了解吗？
- 您会彻底地执行数据分析、数据挖掘所得出的结果吗？

## 数据分析的应用

数据在我们的生活和工作中无处不在，分析工作在各行各业中也变得越来越普及，越来越重要。金融分析家们使用 SPSS 分析工具进行信用卡欺诈检测、信用风险管理；通信业利用 SPSS 分析工具和方法论进行客户管理、产品管理和经营分析；公安部门利用 SPSS 工具进行犯罪监控、物证鉴定；医疗卫生利用 SPSS 工具进行疾病预防控制和不良反应监测；税务机关则可以进行税源管理分析；制造业则可以利用 SPSS 工具进行产品研发设计、质量管理和生产管理；另外，SPSS 分析工具还可以进行个性化的客户关系管理等等。

### 客户分析

我们早已过了供小于求的时代，物资的丰富以及信息的大量共享，使得以产品为中心的 4P 营销理论时代被客户为导向 4C 所替代，而随着网络以及社会化媒体的出现，客户更注重双向交流和关怀等综合性的服务，4R 的营销理论被提上了日程。

顾客是企业利润的来源，所以 CRM 一直是各公司关注的焦点。随着新兴科技和工具的出现，CRM 本身也需要用新的工具和方法论来武装自己。不论企业规模大小，不论企业成熟与

否，我相信在您的企业中肯定会存在以下的一些有关客户的疑虑：如何寻找目标客户群？如何吸引目标客户购买？如何让客户二次购买？如何增加忠诚客户数？如何知道客户何时流失并挽留？如何激发客户的潜在需求并依此开发新产品？如何在正确的时间，推荐正确的产品给正确的客户群体？如果您公司数据库里有客户信息数据、客户交易数据、客户行为数据等，数据分析师就可以利用 SPSS 统计分析工具对客户分群细分，针对不同的群体进行不同的市场活动方案，或针对同一分组的客户进行交叉销售和向上销售；还可以进行 RFM 分析，评估客户终身价值判断客户的生命周期，从而选择正确的时间，推荐正确的商品，达到精准化营销的目的……等等。有效的客户分析可以提高精准营销的准确率，从而降低市场费用，节省了人力和时间，为公司赢得 ROI 的最优化。

### 市场分析

市场分析贯穿了经济活动的整个过程，从生产者到消费者分析，以及各阶段的市场营销活动。这三者是相互影响，不可切割的，生产者分析和消费者分析相对成熟，但是市场营销成本在公司成本中占很大的比重，旧的营销方法是否已经过时？新的营销手段是否适合呢？市场营销是否给销售额带来了贡献呢？

AsiaAnalytics 的营销组合模型 (MMM, Marketing Mix Modeling) 融合了多种统计技术，利用多元回归等统计分析技术分析销售和营销的时间序列数据来估计各种营销组合对销售的影响，以此来优化广告和促销策略方面，以提高销售收入或利润。运用均数差异分析，检验营销要素不同水平之间的差异，方便企业评估各营

- 4P: Product, Price, Place, Promotion
- 4C: Consumer, Cost, Convenience, Communication
- 4R: Reaction, Relativity, Relationship, Retribution
- CRM: Customer Relationship Management
- RFM: Recency, Frequency, Monetary



销因素的作用。利用多因素方差分析、多元方差分析或重复测量及混合效应等方法帮助客户分析和建立优质的营销组合模型。结构方程用图表展示了各因素对销售的影响。

### 销售需求预测

没有一家企业可以不对预测而只是等到事情发生时再采取行动，一个好的短期或长期的经营规划取决于对公司需求的预测。只有正确的预测企业未来的需求，企业才可以科学地分配资源，合理地降低成本。预测不是一件容易实现的事情，它不仅受自身因素的影响，还受到众多外界因素的干扰。过去依赖预测者的经验/主观判断，现在企业依赖统计技术来准确地预测自己产品的市场占有率。

SPSS 的时间序列分析、回归分析能对公司现存历史数据、经济环境、行业变化因素以及经验值进行综合考虑，找出影响销售的主要影响因子，发挥正面影响因子的积极作用，并及时地控制负面影响因子预测未来相关变量的走势……等等。

### 信用风险

信用风险是金融市场中最古老，也是最重要的风险形式之一，现代经济体（特别是金融机构）所面临的主要风险。信用风险管理存在难以量化分析和衡量的问题，并且数据匮乏。

针对信用数据的异质性、高度非线性、数据有限等特点，SPSS 软件可以提供数据模拟技术，在有限数据量的条件下，同样得到准确的结果；聚类分析、异常检测，可以将异质的数据识别出来；Logistic 回归、决策树、神经网络、支持向量机等模型可以很好地解决非线性关系的数据，从众多各种类型的输入变量中，找出影响信用的关键因素，从而预测客户的信用违约概率。量化的信用风险分析帮助金融机构充分了解客户的购买行为和确定正确的信用风险策略，从而提高其市场占有率。

### 欺诈检测

在金融、保险、医疗等行业中，欺诈

一直是让人头疼的问题。欺诈行为的产生通常是因为制度流程的漏洞、信息缺失或不对称造成的，但随着技术的发展，欺诈手段也在不断的发展，这就要求欺诈检测技术不断发展和更新。

SPSS 利用分类预测 (Predictive Modeling)、聚类分析、关联分析 (Association Analysis) 以及异常检测 (Anomaly Detection) 可以从海量的数据中找到异常数据，从而发现客户的异常行为，识别欺诈模式，并利用挖掘出的欺诈模式对新客户行为进行分析并判断；当实际欺诈行为发生后，收集新的欺诈数据，重新对原欺诈模型进行调整，产生能识别新的欺诈行为的欺诈模式。

### 预测维护

应该间隔多久做一次设备的维护？哪些生产环节的因素造成了不合格产品的诞生？一个管理完善的维护维修计划是提高管理效率和设备优化的保障。在进行设备维护时，企业可以选择不同的维护方式，包括事后维护，预防维护，生产停机维护和预测性维护。预测性维护是指根据历史故障的数据来预测未来某一段时间内可能会发生故障的部分，并且能够找出导致故障发生的关键因素。当设备的潜在问题还未到成熟阶段时就被发现，便于提前做好检修准备，从而从容地面对问题的到来，降低突发性问题所带来的损失。因此，相较于其他的维护方式，预测性维护具有维护工作量较少、劳动强度低、诊断维护的准确性高等特点，通过并进行控制关键因素，延长了设备的使用寿命，提高了设备的利用率，减少了维修费用，降低了生产成本，从而提高企业的综合竞争力。

SPSS 提供的预测分析、关联分析、异常检测分析，都可以针对性、准确地提示企业如何维护自己的设备。

### 商业智能

大数据时代的到来，促使了 BI 与预测分析的结合。预测分析的目的是利用历史数据更好地分析未来的行为模式，它是帮助企业优化现有的每一个业务

- 时间序列分析：Time Series 对沿一个方向演化形成的数据序列特征的统计分析。
- 回归分析：Regression 研究一个随机变量 Y 对另一个 (X) 或一组 (X1, X2, ..., Xk) 变量的相依关系的统计分析方法。
- 决策树：Decision Tree, 是在已知各种情况发生概率的基础上，通过构成决策树来求取净现值的期望值大于等于零的概率，评价项目风险，判断其可行性的决策分析方法，是直观运用概率分析的一种图解法。
- 聚类分析：Association Analysis, 是指如果两个或多个事物之间存在一定的关联，那么其中一个事物就能通过其他事物进行预测。它的目的是为了挖掘隐藏在数据间的相互关系。

交易，是一个具有预测性和实时性的工具；BI 是利用现有的历史数据来分析目前的业务状况，以及设定预期的业务目标，它是帮助企业对现有业务进行战略决策的重要工具。

而 BI 和预测分析是非常互补、而又不同的技术，但都是用于决策过程。大数据时代的到来，对于 BI 和预测分析来说，更是如虎添翼。

## SPSS 家族产品介绍

SPSS 家族产品覆盖了整个分析流程，能轻松地访问多种数据类型以及多种集成模块，为分析人员提供任务所需的各种功能，同时该软件的报告和部署功能，可使用户快速清晰地交流结果。下图介绍了 SPSS 家族产品在数据作业流程中所担任的角色。接下来，我们将按照流程，给您介绍 SPSS 家族的各系列产品。



### IBM SPSS Data Collection 数据收集类产品

SPSS Data Collection 是全球应用最广泛的市场调研解决方案，是一个独立界面的集问卷设计、调查实施、数据收集和 data 管理于一体的综合系统，为整个调查过程提供了一个安全、集中的管理平台。该系统涵盖了网络、电话、面访等所有调研方式，执行从问卷设计 - 客户调研 - 数据储存 - 报

告发布的所有过程，支持所有执行方式的不同项目的顾客意见收集，并生成和发布基础数据报告。该系统还可以和 SPSS 统计分析工具，以及和呼叫中心的其他系统同时作业。

#### 主要功能

- 问卷调查：调查量表设计、调查方法选择、避免错答和无回答，从而提高正确回答率。

- 调查执行 / 数据收集：可将在线、邮件、电话、面访等各种调查方式收集到的结果综合成一个数据。
- 调查结果分析和发布：可以基于 Web 发布调查结果的数据统计报告。
- 调查结果报告和输出：不仅可以基于 Web 发布，还可以与 SPSS 或 Office 产品进行结合，可以很容易地生成有关格式报告。

### IBM SPSS Statistics 统计分析类产品

世界上最早采用图形菜单驱动界面的统计软件，最突出的特点就是操作界面极为友好，输出结果美观漂亮。几乎所有的功能都能以统一、规范的界面展现出来，能使用 Windows 窗口的方式展示各种管理和分析数据方法，对话框展示出各种功能选择项。SPSS 采用类似 EXCEL 表格的方式输入与管理数据，数据接口十分通用，能方便地从其它数据库中读入数据。其统计过程包括了常用的、较为成熟的统计过程，完全可以满足专业人士的工作需要。输出结果十分美观，存储时则是专用的 SPV 格式，可以转存为 HTML 格式、纯文本格式、微软 Word、PowerPoint、PDF 文档。

对于熟悉编程运行方式的用户，SPSS 还特别设计了语法生成窗口，用户只需在菜单中选择相关的选项，然后按“粘贴”按钮就可以自动生成标准的 SPSS 程序，极大地方便了中、高级用户。

### IBM SPSS Modeler 数据挖掘类产品

作为一个企业级数据挖掘平台，Modeler 结合商业技术可以快速建立预测性模型，进而应用到商业活动中，帮助人们改进决策过程。Modeler 数据挖掘工具被广泛应用于数据挖掘的整个流程，帮助数据挖掘和分析人员快速访问并准备建模所需的数据，建立起模型并进行拟合性检测和比较，并能立刻将模型应用于实际环境。决策者可以将 Modeler 无缝地整合到其他商业流

程和系统中。Modeler 可以导出所有的数据挖掘步骤，包括数据访问、建模和加工。

#### 主要功能

- 界面简单直观：通过分析节点和箭头形成流。
- 易学、易用、专业性：涵盖了最先进的机器学习、人工智能、模式识别和统计学建模方法。
- 快速发现最优模型、自动数据处理和建模：可以比较在同一过程中不同模型的准确率、提升率、响应速度。
- 团队协作平台：界面的直观化方便分析人员与用户进行更好的交流。
- 灵活性、可扩展性：开发的数据库接口几乎支持全部的关系型数据库。
- 更广泛的数据来源：补充 SPSS 其它产品功能后，Modeler 能整合结构化和非结构化数据来源。
- 企业级标准和技术：连接各种应用系统平台里的多种样式的数据库、数据表和无格式文本文件。

### IBM SPSS C&DS (Collaboration and Deployment Services) 结果发布类产品

提供了企业级的分析资产管理和分析功能部署平台。目的在于对 SPSS Statistics、SPSS Modeler 等产品分析结果进行统一管理和部署。促进组织分析人员的协同合作，使分析结果的发布更快捷有效；自动化分析过程提高了分析结果的一致性和可控性；分析结果自动分发给相关人员和系统。

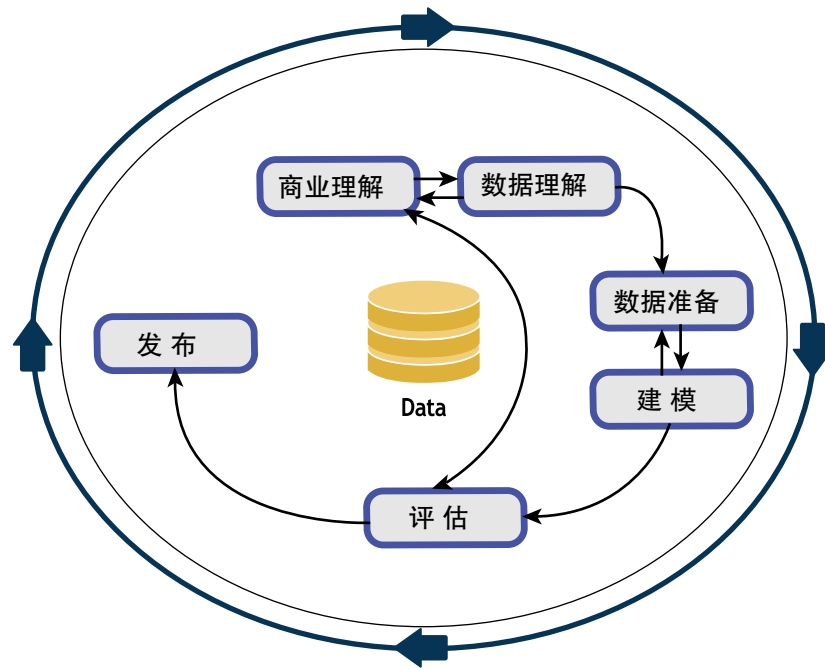
#### 主要功能

- 协作：管理分析资源，共享分析结果，提高效率。
- 自动化：提供可靠的分析流程自动化处理。
- 部署：在企业内更容易地部署分析结果，更好地支持决策。

- 机器学习 (Machine Learning) 是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。
- 人工智能 (Artificial Intelligence) 是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

## 统计分析和数据挖掘方法论

CRISP-DM 是业界公认的数据挖掘方法论，其完善的数据挖掘项目管理过程，可以从业务理解到结果发布的全过程对数据挖掘项目进行有效地管理。它包含了各个典型阶段的说明、每个阶段所包含的任务以及这些任务之间的关系说明，CRISP-DM 概述了数据挖掘的生命周期。



### 商业理解 (Business Understanding)

- 确定业务目的 · 状况评估 · 数据挖掘目的评估 · 立项

### 数据理解 (Data Understanding)

- 收集初期数据 · 数据描述 · 数据探索 · 验证数据质量

### 数据准备 (Data Preparation)

- 数据抽取 · 数据选择 · 数据清洗 · 数据生成 · 数据合并 · 设定数据类型

### 建模 (Modeling)

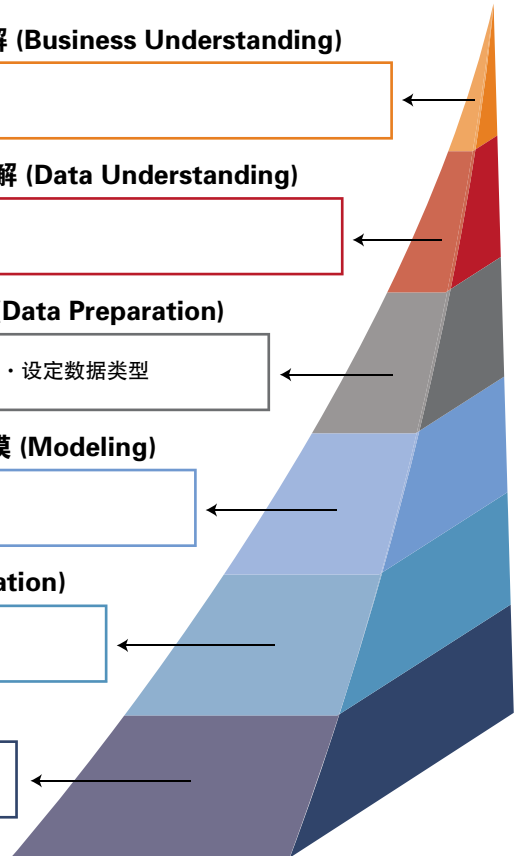
- 选择建模方法 · 生成测试设置 · 模型创建 · 模型评估

### 评估 (Evaluation)

- 结果评估 · 再次检查过程 · 检查后续阶段

### 发布 (Deployment)

- 确立发布计划 · 确立控制/维护计划
- 制作最终报告书 · 再次检查项目





SPSS 公司成立于 1968 年，总部位于芝加哥，自 2009 年 10 月加入 IBM 旗下。在纳斯达克上市的股票代码为 SPSS。SPSS China 成立于 2006 年 5 月 18 日，总部位于上海，在北京设有分公司。SPSS Korea, SPSS China, SPSS Taiwan 同属于东亚团队，三地一体化的运营管理模式，所以自 2013 年 1 月 1 日起，统一更换英文名为“AsiaAnalytics”。

SPSS 公司在全球约有 30 多万家产品用户，它们分布于通讯、银行金融、保险证券、制造业、市场调研、政府机构、教育科研、医疗卫生、化工行业、零售连锁、电子商务等多个领域和行业，全球 500 强中约有 80% 的公司使用 SPSS，而在市场研究和市场调查领域有超过 90% 的市场占有率，是世界上应用最广泛的专业统计软件之一！SPSS 在财政金融、政府机构、教育、电信、市场研究、零售、电子商务等领域已有 40 多年的数据分析经验，提供从数据输入、数据整理、探索分析、分析报告、建立模型、预测分析到结果发布的完整解决方案，使用户能够更好地预测未来，把握先机。

析数软件(上海)有限公司  
AsiaAnalytics 上海  
电话: +86 21 - 6352 3300  
传真: +86 21 - 6352 3511  
地址: 上海市延安东路 550 号  
海洋大厦 905 室

AsiaAnalytics 北京  
电话: +86 10 - 8523 3290  
地址: 北京市朝阳区建国门外大街  
甲 12 号 新华保险大厦 15 层 1573 室





AsiaAnalytics

Data. Insights. Performance.

Business  
Partner



TM