# 0617重新认识数据分析

1. 内容介绍
   1. 学习目标
      1. 对数据分析建立起**系统、全面**且**正确**的认知
      2. 理清且理解数据分析领域的各种专业术语
   2. 解决的问题
      1. 数据是啥？数据分析是啥？它们有什么用？等概念不清
      2. 如何让数据在个人和企业里发挥最大价值？
   3. 知识点
      1. 数据自身的三大属性：直击数据本质，视角高屋建瓴
      2. 数据建设的七个层级：理清发展逻辑，建设商业智能
      3. 数据分析的实际价值：全面认知价值，提升能力上限
      4. 数据分析的实现原理：知晓实现原理，扫清知识盲区
      5. 数据分析的能力范围：明确能力范围，做到有的放矢
      6. 数据分析的职业发展：摸清职业生态，指明职业方向
2. 具体内容
   1. 两个问题
      1. 第一个问题：经营一家只能坐5个人的麻辣烫街边铺，需不需要用数据分析？
         1. 不需要
      2. 第二个问题：经营一家麻辣烫上市公司，需不需要用数据分析？
         1. 肯定需要
   2. 从不需要数据分析到一定需要数据分析，中间到底发生了什么？
      1. 故事从一名有梦想的麻辣烫老板说起，而你就是这位老板
      2. 凭感觉
         1. 需要考虑哪些问题
         2. 考虑问题的先后顺序
         3. 按照特定逻辑的拆分和列举，已经让关键问题变得清晰起来
         4. 接下来就是如何解决这些问题
      3. 看数字 + 用表格
         1. 其实，之前的问题分别对应了如今智慧餐饮领域的关键分析方向
            1. 多少钱？ --- 财务/现金流管理及分析
            2. 卖什么？ --- 选品分析
            3. 怎么做？ --- 配方/制作工艺工业化研发
            4. 开在哪？ --- 智能店铺选址
            5. 哪进货？ --- 供应链管理
            6. 咋获客？ --- 线上流量运营
            7. 咋复购？ --- 用户/会员系统与分析
            8. 咋持续？ --- 经营分析
         2. 选址可是一家线下店铺成功的关键因素，我们就以【开在哪】这个问题举例
            1. 看数字

至少统计出各个备用选址在用餐时间的平均人流量，以及几天内的极大值和极小值

麦肯锡等知名咨询公司，大项目都是要到线下数人流的

平均数、极大值和极小值这些基础的描述统计学知识上线了

而我们看数字时，数据自身的两大优势属性就体现出来了

客观

用数字衡量和表现一件客观事物时，能最大程度统一大家的认知

量化

可以进行各种精密的计算，得出更有价值的结果

* + - * 1. 单纯的计算一些简单的数字肯定不够

用表格

* + 1. 搭系统
       1. 再利用上表格数据后，我们已经能解决一家门店大多数的业务分析问题了
          1. 多少钱？ --- 财务/现金流管理及分析

用Excel测算财务模型

* + - * 1. 卖什么？ --- 选品分析

去竞品家统计食材消耗

根据自己的食材剩余优化

* + - * 1. 怎么做？ --- 配方/制作工艺工业化研发

每次制作严格按照量化的配方和时间进行

* + - * 1. 开在哪？ --- 智能店铺选址

汇总整理选址数据，量化进行分析

* + - * 1. 哪进货？ --- 供应链管理

根据前几天的食材消耗订次日的食材

每天盘点食材库存并登记

* + - * 1. 咋获客？ --- 线上流量运营

入住大众点评/饿了么等平台增加曝光

从平台后台下载数据表格进行分析

* + - 1. 但是随着业务规模的扩大，品牌需要扩展，此时如何对多家门店的数据进行统计
         1. 人工记录
         2. 开发一整套信息服务系统--平时所见的POS机或收款机
         3. 在搭建完企业自身的信息服务系统后，一个可规模化拓展能满足各种业务所需求的商业模式就初步形成了！
    1. 建数仓
       1. 随着数据系统的不断完善，每日全国各地门店的数据汇总已经超过Excel的储存上限
       2. 企业需要建设自己更大的数据仓库，处理不断增长的数据资产
       3. 而数据仓库技术占到数据分析实现所用技术的70%以上
          1. **示意图**



* + - * 1. 埋点采集（数仓）

传感器

系统日志采集方法

Web爬虫

* + - * 1. 传输存储（数仓）

传输

借助网络

存储

借助数据库技术

研究如何组织和存储数据，如何高效的获取和处理数据

存储是门大学问

* + - * 1. 提取回流（数仓/算法）

提取

查询

导出

回流

数据被操作过后，怎么回到数据库，继续参与生产和建模

* + - * 1. 清洗处理（数仓）

清洗

符合数据标准

处理

符合实际需求

* + - * 1. 连接导入（数仓）

怎么对接到不同的工具和软件，涉及到各种中间键和协议

* + - * 1. 分析建模（分析/算法）

分析

计算

图表

建模

机器学习

深度学习

* + - * 1. 部署应用（数仓/分析/算法/开发/产品）

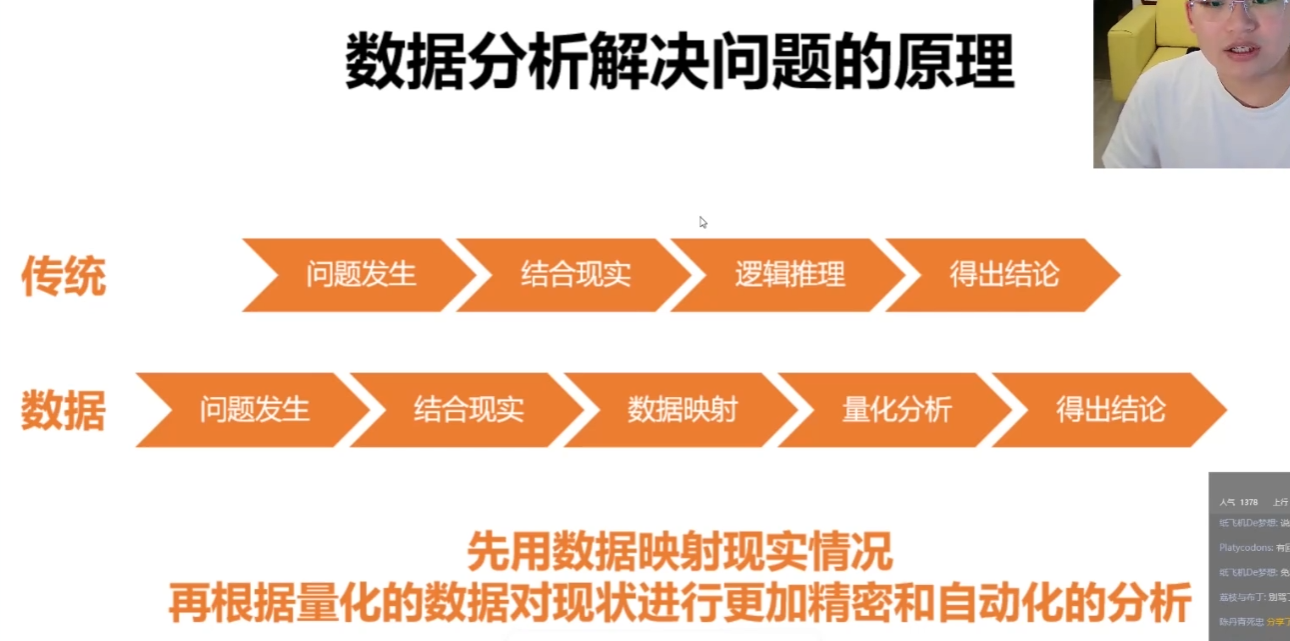
BI 监控

算法平台

先知

工程开发

* + - 1. 在数据仓库完全准备完毕后，我们的麻辣烫公司招募了第一名数据分析师。那么他又是如何基于数据分析的问题呢？
         1. 示意图



* + - * 1. 业务中的数据分析是为了更好地解决现实问题

传统：问题发生 -- 结合现实 -- 逻辑推理 -- 得出结论

数据：问题发生 -- 结合现实 -- 数据映射 -- 量化分析 -- 得出结论

* + - * 1. 数据分析只是分析的一种，没有数据一样要进行分析

能对碎片化的信息加以抽象和整合，不过度依赖量化，才是真正的分析能力

数字数据，经验数据

* + - * 1. 数据分析实现的具体流程

示意图



定义问题

明确范围

拆解问题

准备数据

获取数据

清洗数据

处理数据

验证数据

分析数据

分析的目的

业务分析：关注数据和业务的关联性，映射关系和可解释性

常规计算

数据可视化

归纳演绎

思维模型

建模分析：关注数据的整体特征（特征决定上限，模型调参无限接近于上限）

机器学习建模

深度学习建模

分析的种类

计算型分析

统计学

机器学习

逻辑性分析

归纳演绎

思维模型

得出结论

假设性结论

初步传递结论

得到新的反馈

验证/推翻结论

应用数据

策略 + 实施：业务解决问题

监控 + 预警：自动发现问题

算法 + 部署：自动解决问题

* + 1. 中心化
       1. 当数据仓库建立完毕后，企业的数据团队基本上也搭建完毕了，但初级的数仓存在问题
          1. 跨库查询：数据存储在不同服务器中，无法实现业务级的跨库查询
          2. 数据架构：数据由每个系统的开发独立维护，指标定义、数据结构、数据类型等架构标准都不统一
          3. 数据需求：字段需求无法精准有节奏地一一实现，开发说不做就不做了
       2. 中心化的数据库将会统一管理企业的所有数据字段，将所有数据汇总到一个技术架构和仓库里
       3. 只有一线C端的大型互联网公司才完成了数据中心的建设
          1. 京东B轮额时候才进行了数仓的全面重构和中心化
          2. 淘宝也是在阿里云出世后才彻底完成
          3. 字节很多新增业务部依旧没法接入数据中心

（接入不够敏捷，还没接入就凉了，于是数仓就会拖需求，等你凉）

* + - * 1. 师兄有幸在第三份工作里才用上了中心化的数据库
    1. 智能化
       1. 在完成数据中心的建设后，企业在数据征途上将挑战一个永无止境目标 -- 商业智能
       2. 商业智能实现的四个阶段
          1. 示意图



* + - * 1. 反映现实情况

指标体系

数据挖掘

* + - * 1. 监控关键指标
        2. 辅助决策制定

分析建模

数据实验

* + - * 1. 实现自动决策

智能算法

工程部署

真正实现智能化自动化

* 1. 我们应该如何在需要数据分析的企业里发挥价值？
     1. 数据分析对企业、个人乃至时代到底有什么价值？（数据分析的实际价值）
        1. **价值即上限**
        2. 个人价值
           1. 表面上

有效提升职场竞争力

通用、万精油涨薪神器

未来高级岗位的必备技能

* + - * 1. 深入思考一层

为什么你会数据分析，公司就愿意高价聘请你，你不会公司就不愿意高薪用呢？

数据分析

是帮公司发现商机？

是帮公司解决问题？

* + - * 1. 我们来对比一下会数据分析和不会数据分析的员工在公司中工作的实际场景

发现今天的销售额/业绩突然下降了，想弄清楚原因

想要复盘一下近期的工作成果

开会

数据量化

想要发现新的市场机会

传统：跑市场、拉关系

数据：宏观数据、舆情分析、研究报告

想要扩大业务规模

传统：招人、扩充团队

数据：提高软件服务效率、自动化、增加服务器、建模

* + - * 1. 对比完这些场景，你认为数据对个人最大的价值是什么？

提升工作效率

数据专员做表也是提高效率，因为业务自己弄太慢

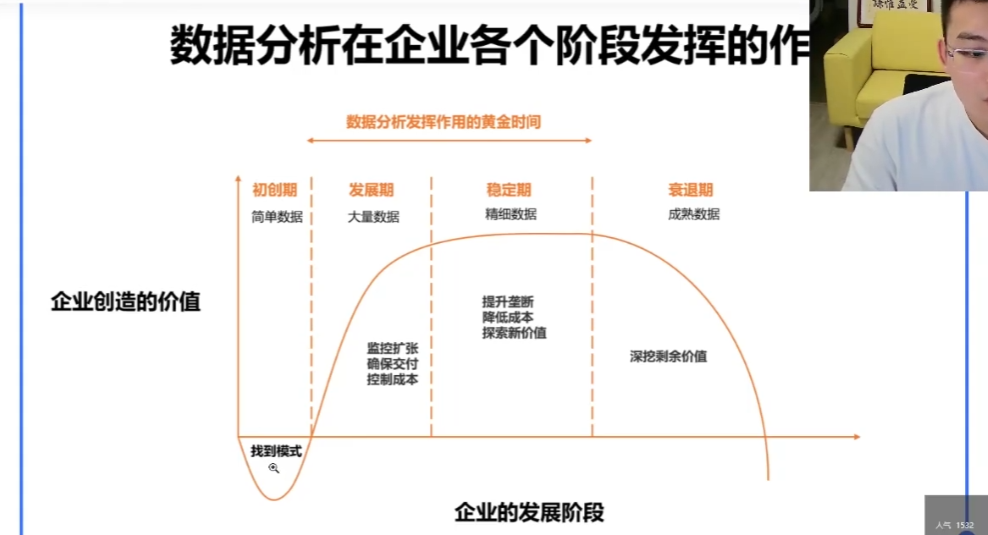
搭监控也是为了提高效率，可以随便看

算法工程师建模也是提高效率，让机器可以自动处理信息

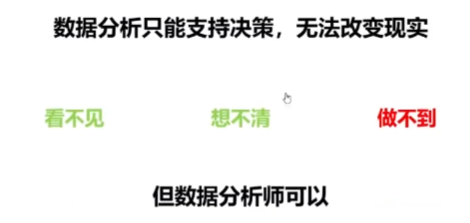
师兄作图也是提高效率，为了让观点更加高效的传达

**对数据型人才的需求其实是对高效率人才的需求** -- **效率是数据分析师的底线也是生命线**

* + - * 1. 如果说数据对个人最大的价值是体现在效率提升上，那数据对公司的价值又体现在哪些方面呢？
      1. 公司价值
         1. 映射现状，及时**反馈**
         2. 统一认知，加强**协作**
         3. 算法基础，自动**生产**
      2. 时代价值
         1. 数据是一切科学研究的基础
         2. 科学研究就是：官场世界 -- 提出假设 -- 设计实验 -- 得到数据 -- 验证假设 -- 得出规律
      3. **价值即上限**
    1. 企业在什么时候会需要数据分析？（数据分析的能力范围）
       1. 企业发展阶段
          1. 示意图



* + - * 1. 初创期 - 简单数据 -- 找到模式
        2. **发展期 - 大量数据 -- 监控扩张、确保交付、控制成本**
        3. **稳定期 - 精细数据 -- 提升垄断、降低成本、探索新价值**
        4. 衰退期 - 成熟数据 -- 深挖剩余价值
      1. 企业问题类型
         1. 示意图

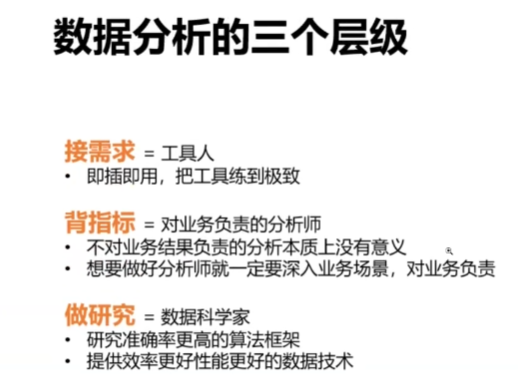


* + - * 1. 看不见 - 数据探索
        2. 想不清 - 数据分析 - 做规划
        3. 做不到 - 对不起，这个只能靠实践和执行力，靠资源靠成本

本质不是分析没用，而是这件事本身没用或者做不成

数据分析只能支持决策，无法改变事实，但分析师可以。而你会做出怎样的选择？

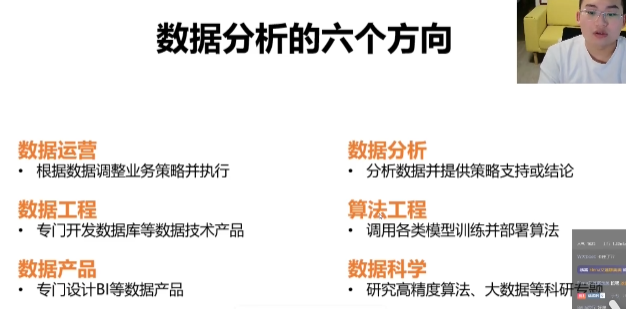
* + 1. 企业需要哪些种类的数据分析？（数据分析的职业发展）
       1. 三个层级



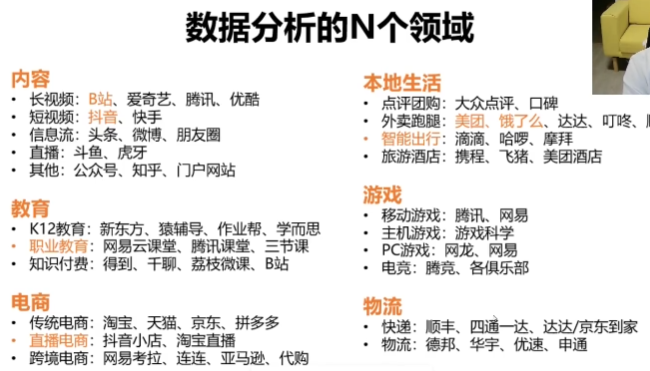
* + - 1. 四种类型



* + - 1. 六个方向



* + - 1. N个领域



1. 知识点总结
   1. 数据自身的三大属性：直击数据本质，视角高屋建瓴
   2. 数据建设的七个层级：理清发展逻辑，建设商业智能
   3. 数据分析的实际价值：全面认知价值，提升能力上限
   4. 数据分析的实现原理：知晓实现原理，扫清知识盲区
   5. 数据分析的能力范围：明确能力范围，做到有的放矢
   6. 数据分析的职业发展：摸清职业生态，指明职业方向
2. 知识点自检清单
   1. 数据自身的三大属性：直接数据本质，视角高屋建瓴
      1. 数据自身有哪些属性？
         1. 客观
         2. 可量化
      2. 这些属性有什么用？
   2. 数据建设的七个层级：理清发展逻辑，建设商业智能
      1. 企业的数据建设要经历哪些层级/阶段？
      2. 各个层级的区别是什么？
      3. 为什么高层级的数据建设能更好地支持数据分析的工作？
      4. 你所待过的公司在那个层级
   3. 数据分析的实际价值：全面认知价值，提升能力上限
      1. 数据分析对个人的最大价值是什么？
      2. 数据分析是如何帮公司创造价值的？
      3. 为什么数据分析是我们这个时代科学技术进步的关键？
   4. 数据分析的实现原理：知晓实现原理，扫清知识盲区
      1. 用数据解决问题和拍脑袋解决问题有什么不同？
      2. 数据分析流程中第一步和最后两步是什么？
      3. 数据分析初步得出的结论可以直接用于实际业务吗？
      4. 根据分析目的的不同，有几种分析？它们各自的特征是什么？
      5. 根据分析种类的不同，有几种分析，它们各自的特征是什么？
      6. 清洗数据后需不需要再做验证和确认？
   5. 数据分析的能力范围：明确能力范围，做到有的放矢
      1. 企业在哪几个阶段最需要数据分析？
      2. 数据分析能帮企业解决怎样的问题？
      3. 数据分析补鞥呢帮企业解决怎样的问题？
      4. 分析师面对难以解决的问题时应该怎么做？
   6. 数据分析的职业发展：摸清职业生态，指明职业方向
      1. 背指标的分析师和接需求的分析师有什么区别？他们长期的薪资发展谁更好？
      2. 四种类型的数据分析各自的功能以及负责的核心指标有什么不同？你更喜欢哪一个？
      3. 数据分析职业发展共有几个方向？这几个方向的区别在哪？你更喜欢哪一个？
      4. 数据分析最热门的三个领域是哪三个》你最感兴趣的领域有哪些？
3. 课后练习
   1. 总结你对数据分析的认知在课程前后有什么不同和偏差，并完成知识自检清单
   2. 以幕布分享链接的形式上传到评论区
   3. 幕布格式
      1. 对数据分析的认知
         1. 上课前
         2. 上课后
         3. 最大的不同
4. 学习的困难只有两个
   1. 找到正确的路径
   2. 攻克单点的难关