

## 一 数据科学领域概述

一位老前辈说：选对方向，努力才有意义。刚毕业的薪资可能跟学历有关，但有的行业天花板很低，直接就决定了你的薪资上限，再努力都很难突破。而IT互联网行业的薪资天花板就高的多。究竟能拿多少，完全看个人了，从最低月薪到年薪几十万。IT数据分析岗位，市场需求越来越大。

<https://index.baidu.com/v2/main/index.html#/crowd/%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90?wolds=%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90.%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0.%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0.%E8%BD%AF%E4%BB%B6%E5%BC%80%E5%8F%91>

### 1.1 数据如何驱动运营给企业带来价值

- 原始收集数据
  - 数据埋点 收集用户在网页端，APP，小程序等终端的各种数据
  - 业务数据
  - 外部数据
- 数据加工处理
  - 将收集的数据转换成可理解、可量化、可观察的业务指标
  - 单纯的数据没有意义，只有和业务结合才能发挥价值
- 数据可视化
  - 有了数据指标，必须管理好指标
  - 建立指标体系
- 数据决策和执行
  - 从数据中得到相关信息，需要把这些信息转换成策略
  - 包括策略制定，并持续优化和改进策略
- 数据产品
  - 将策略制作成数据应用和产品
  - 开始自动化和系统化运营
- 数据战略
  - 积累了大量数据，大量模型，大量数据应用
  - 不只是数据分析，可以将数据变现

### 1.2 数据科学落地架构



- 三类岗位
  - 业务线 数据分析
  - 研发线 数据仓库
  - 算法线 数据挖掘
- 关键词说明
  - ETL 清洗 转换 加载
  - DW 数据仓库
  - CRM 客户关系管理
  - CMS:Content Management System "内容管理系统"
- 不同岗位对业务知识的要求

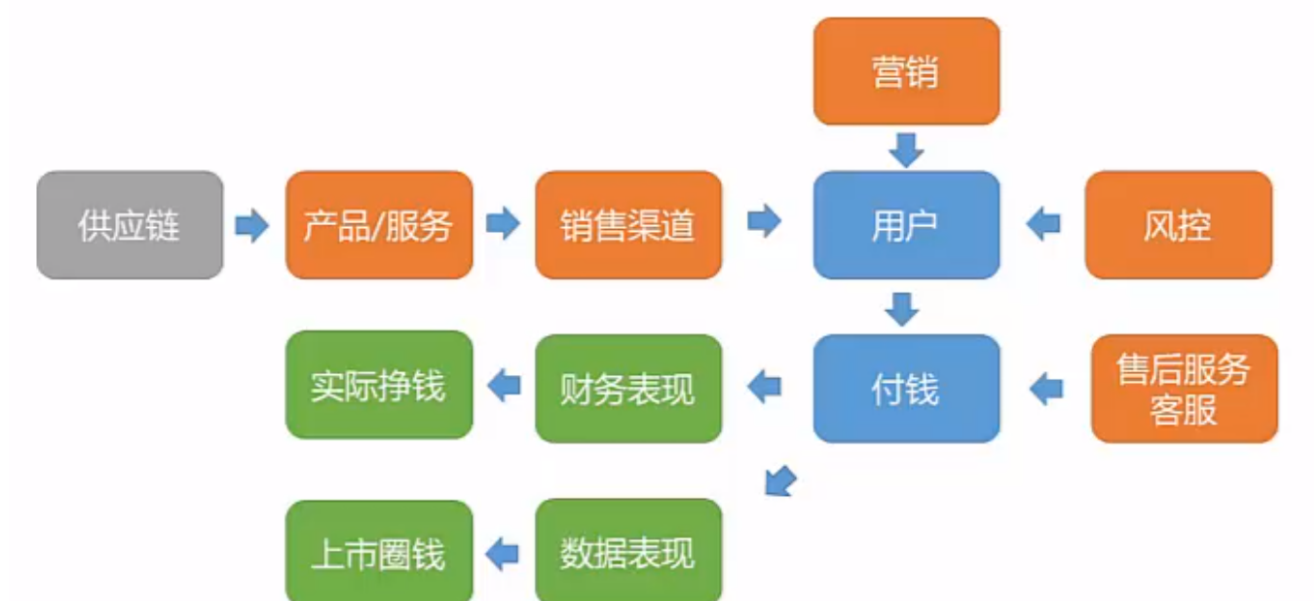
	了解基本概念	知道指标含义	熟悉分析套路	掌握分析思维	独立处理问题
业务分析师					√
数据运营					√
数据分析师				√	
数据挖掘工程师				√	
算法工程师				√	
数仓工程师		√			

## 二 什么是业务

“业务”更白话一些来说，就是各行业中需要处理的事务，但通常偏向指销售的事务，因为任何公司单位最终仍然是以销售产品、销售服务、销售技术等等为主。“业务”最终的目的是“售出产品，换取利润”。所以通常会把业务员等于销售员，也就是这个原因，业务就是进行或处理商业上相关的活动。业务也是渠道就是指产家与销点之间关系是通过渠道之间建立起来的。而业务员在这中间起了重大的作用。业务员的工作直接影响到产家、销点、消费者这三者之间的关系。”

[业务\\_百度百科](#)  
[baike.baidu.com](http://baike.baidu.com)

## 2.1 理解业务从理解商业模式开始



- 商业模式（怎么挣钱）
  - B2C
  - B2B
  - B2B2C
  - B2VC:
- 不同类型企业分类

## 互联网企业：虚拟产品、B2VC

- ① 游戏：B2C（T、网易、完美、37）
- ② 电商平台：B2B（A、京东、拼多多）
- ③ 广告：B2B（B、A、T、头条、360）
- ④ 直营电商：B2C（网易严选、京东自营）
- ⑤ 社交：B2B+B2C（T、探探）
- ⑥ 互金：B2C（互联网支付、消费贷、P2P）
- ⑦ O2O平台：B2B+B2C（美团、饿了么、滴滴）
- ⑧ 新闻资讯：B2B+B2C

- 理解商业模式的关键问题

- 业务模式
- 四大角色
  - 产品 卖的是什么服务
  - 渠道 以何种方式联系到客户
  - 用户 最终消费者 直接使用者
  - 运营 营销、服务、风控手段
- 部门 自己要面试/服务的对象
- KPI（Key Performance Indicator）

- 数据科学在商业行为中的角色：解决产品 销售 运营 营销等数据问题，驱动达成商业目标

## 实体企业：实际产品、毛利

- ① 零售：B2C（屈臣氏、大悦城）
- ② 快消：B2C（宝洁、雀巢）
- ③ 母婴：B2C（合生元）
- ④ 耐用：B2B（美的、各大车厂）
- ⑤ 通讯：B2C
- ⑥ 经销商：B2B2C
- ⑦ 资讯服务：B2B（广告公司、咨询公司）
- ⑧ 开发服务：B2B（美团、饿了么）

## 三 数据分析工作需求

# 数据分析师 20-21K

首汽约车 [查看所有职位](#)

## 工作职责：

- 1.负责风控业务数据的采集、整理、分析和维护
- 2.分析与发现业务环节中的风控点，指定风控规则，有效止损
- 3.数据挖掘分析，风险数据监测与解读，发现风险点，提出改进性建议
- 4.推动技术部门与完善建立风险数据分析模型

## 职位要求：

- 1.统计学、数学、计算机相关业务或工作背景优先
- 2.对数据敏感，较强的逻辑分析能力和文字表达能力
- 3.熟练的Excel操作技能，掌握R，SPSS，SAS和Python的优先
- 4.熟练的数据库优先，掌握hive sql优先
- 5.熟悉数据挖掘的思路和常用的数据挖掘算法
- 6.良好的沟通能力，有责任心，学习能力强，能够承担一定的工作压力
- 7.具有互联网公司数据分析经验的优先

# 数据分析师 10-15K

五险一金

补充医疗保险

定期体检

年终奖

带薪年假

...

中信银行 [查看所有职位](#)

社招要求:

- 1、年龄33岁以下，全日制本科及以上学历，数学、统计、计算机、金融类专业；
- 2、熟悉SQL大数据量数据处理开发，具备数据仓库、数据集市等数据应用项目实施经验，或核心等主要业务系统后台数据库开发经验，对主要银行业务有一定的了解；
- 3、在机器学习、数据挖掘、图像、自然语言理解等一个或多个专业领域有扎实的理论知识优先考虑，熟悉tensorflow、caffe等其中一种深度学习框架，有过相关项目经验优先；
- 4、有扎实的编程基础，至少精通一种编程语言，如PYTHON、R、SAS、C、JAVA优先考虑；
- 5、具有优秀的分析问题和解决问题的能力；
- 6、责任心强、积极主动，有良好的沟通能力和团队合作能力。

# 数据分析师 40-45K

五险一金

补充医疗保险

定期体检

年终奖

带薪年假

...

兴业银行北京分行 [查看所有职位](#)

工作职责：

- 1、对行内现有客户的特点及行为进行分析，提炼客户风险特征；为模型及策略的改进奠定基础；
- 2、负责构建风险监控指标及相应监控手段，确保自己负责的风险在可控范围；
- 3、针对业务风险的变化情况进行分析，找到风险控制的相关策略，并跟进策略的落地情况；
- 4、跟进客户的流程转化率情况，评估流程中客户流失原因的合理性；
- 5、监控整体或部分资产组合的风险情况及风险变化过程，预测未来资产的质量情况。

任职要求：

- 1、全日制硕士及以上学历毕业，2年及以上大量用户数据分析、用户研究工作经验；
- 2、精通sql，excel透视表等，熟练掌握统计学方法与工具；
- 3、逻辑清晰，主动进取，较强的沟通能力，善于团队合作；
- 4、具有较强结构化思维、逻辑思维能力，优异的学习能力，对数据敏感，具备优秀的信息整合和分析能力，能够形成清晰的业务观点和前瞻判断，了解互联网常用的数据分析方法和思路。

# 银行数据分析师 15-25K

五险一金

补充医疗保险

定期体检

年终奖

带薪年假

...

中信百信银行

[查看所有职位](#)

## 职责描述：

负责部门数据建设、分析、使用等方面的整体规划；

负责分析部门内业务指标定位、用户画像；

负责事业部内部的数据模型建设；

负责日常经营数据的统计分析；

负责报表的开发与设计；

## 任职要求：

银行、支付行业分析相关工作5年以上；

熟练使用sql, hive, smartbi等产品；

了解银行账户、支付、清算等业务者优先；



# 数据分析师 15-30K

五险一金

定期体检

带薪年假

免费班车

餐补

...

百度 [查看所有职位](#)

- 1、通过对数据的敏锐洞察以及定性和定量分析，迅速定位内部问题或发现机会
- 2、负责大商业（凤巢&Feed&品牌&展示广告）日常数据分析及监控，针对异常情况协调资源进行跟踪和深入分析
- 3、能根据实际业务完成较深入的专项数据分析，并形成数据分析报告
- 4、负责大商业数据分析体系建设和完善，挖掘流量、产品、策略方面的商业变现机会，驱动大商业业务发展

任职要求：

- 1、统计学、数学、经济学等相关专业，扎实的机器学习或数据挖掘理论和技术基础，具有相关领域较深入的技术应用的经验和能力
- 2、三年以上数据分析经验，至少熟练使用R、SQL、SAS、MATLAB、PYTHON中的一种，熟练掌握SQL，熟悉常用数据统计和分析方法
- 3、良好的沟通能力、团队合作精神、工作规划能力和主动意识
- 4、良好的逻辑思维能力、业务解读能力和快速学习能力，能够独立领导完整的数据分析项目

# 数据分析师 13-26K·13薪

五险一金

补充医疗保险

定期体检

全勤奖

年终奖

...

京东 [查看所有职位](#)

## 岗位描述:

- 1、以提升运营质量为目的，对京东物流现有核心数据进行模型分析
- 2、从数据方面找出指标之间的相互关联关系及运营异常情况
- 3、支持配送各部门提数需求
- 4、完善提数逻辑，明确已经推送的数据取数逻辑表

## 任职要求:

- 1、熟练掌握数据报表处理工具，如：Excel、SQL、Spss等；
- 2、在数据挖掘、分析和建模方面有丰富经验；
- 3、3年以上数据分析经验；
- 4、数学专业或物流行业经验者优先考虑

## 数据分析师 / 10k-20k

☆ 收藏

投个简历

## 职位描述:

### 岗位职责:

- 1、能够主动使用分析方法，结合业务发展需求，提出创造性业务策略，促进业务发展；
- 2、不断整合与挖掘已有策略的运营问题点，创造合理优化建议，调整策略方案内容，完善业务策略；
- 3、数据指标的SQL提取，及策略指标测算工作，保证指标逻辑及阈值合理性；
- 4、研究加盟配送考核策略的特点，结合业务目标，对不同时期及季节提出更有针对性的解决方案。

### 任职资格:

- 1、本科以上，5年及以上工作经验，有O2O行业互联网相关数据分析工作经验者优先；
- 2、具备优秀的数据分析能力，掌握SQL，独立高效完成业务策略；
- 3、具备丰富的数据规则、考核方案、运营策略的经验，从数据中发现痛点并优化策略；
- 4、逻辑清晰，主动进取，较强的沟通能力，善于团队合作，主动思考及推进工作；
- 5、行业视野广，能适应高强度工作环境，抗压能力强。

### 工作地址

北京 - 朝阳区 - 锐创国际大厦B座4层

[查看地图](#)

美团点评

消费生活

上市公司

2000人以上

<http://www.meituan.com>

## 相似职位

轻任

数据分析师

10k-20k

美住 [北京·朝阳区]

数据分析师 / 13k-25k

收藏

投个简历

职位描述:

岗位职责:  
1、搭建用户层级的数据分析通用框架  
2、衔接技术搭建多用户视角、多维度分层视角的数据池  
3、深入理解业务、内容和用户,能通过数据分析三者特点,进行数据反馈,挖掘用户需求。  
  
岗位要求:  
1、国家全日制本科及以上学历,1年以上相关工作经验。  
2、熟练掌握SQL语言或其他数据分析软件,可以用数据分析建立相关模型,能理解业界常用的各类数据分析模型及算法。  
3、具备优秀的信息整合、逻辑组织和观点提炼的能力,能够独立完成数据分析报告并提出建设性见解,有热情,愿意长期从事相关行业市场分析和战略规划工作。  
4、Excel精通,熟练使用Tableau、R、Python等分析软件中的一种。

友可乙

爱奇艺

88 文娱 | 内容

上市公司

2000人以上

http://www.iqiyi.com/

相似职位

数据分析师

13k-25k

数据分析师 / 13k-25k

收藏

投个简历

1) 负责业务产品或算法建模效果数据分析,并根据分析结果提出优化建议,实现数据驱动业务产品或者策略的改动  
2) 建立新浪新闻客户端及相关产品的用户数据体系,包括用户分层,用户转化路径等  
3) 负责撰写专题性数据分析报告,发现产品或业务的问题关键并可通过数据分析和内在联系,为产品决策提供合理化建议;  
4) 负责新浪新闻客户端和手机新浪网核心数据指标和新产品功能的日志收集、指标定义、数据监控、数据异常分析;  
5) 及时响应产品、运营、市场或者BD的数据统计和分析需求;

任职要求:  
1) 统计学或计算机相关专业本科以上学历,3年以上互联网数据分析相关经验。  
2) 熟悉常用的数据建模方法如回归、聚类、关联分析等,熟练操作数据提取、处理和分析软件,如SQL/Hql、Excel、R、SPSS等。  
3) 熟悉Hadoop使用,了解HTML、Javascript、CSS、PHP等页面制作技术。  
4) 对数据敏感、具备较强的数据分析能力、逻辑思考能力、沟通能力。  
5) 具备较强的数据报表、数据报告制作能力。有用户相关分析经验及建模经验优先

上市公司

2000人以上

http://www.sina.com.cn

相似职位

数据分析师

14k-25k

先花小贷 [北京·酒仙桥]

数据分析师

15k-25k

数美 [北京·望京]

数据分析师

18k-25k

小蒂规划 [北京·望京]

数据分析师

15k-25k

猫眼电影 [北京·和平里]

- 数据分析需要掌握的技能
  - 业务知识,业务思维
  - Excel
  - 数据可视化
  - SQL
  - 统计学(机器学习模型)
  - Python

## 四 数据分析日常工作介绍

### 4.1 日、周、月报

- 日报(每天必须要看):
  - 各种数据源整理,了解业务现状
  - 数据波动监控解释
  - 长期观测数据可以培养数据敏感性

- 周报（短期复盘）
  - 短期趋势
    - 一般公司发版频率 1~2周
  - 版本迭代分析
    - 微博 单周发版
  - 为其它内容做积累
- 月报（决策）
  - 每个月都会例会
  - 业务梳理
    - 最近一个月 KPI比较好，因为活动效果比较好
    - 分析原因，同步到各个部门
  - 部分建议
  - 目标评估及战略决策

## 4.2 临时数据 数据需求

- 管理层要数据
- 业务执行人员需求
  - 永无止境，不要立刻就做
  - 为何要这样做
  - 如何落地
- 案例
  - 某APP最近上线了一个线下餐饮优惠券功能，产品经理需要快速知道优惠券的使用人数，你怎么做？
    - plan A 先分析一下字段口径，在哪个是日志中提取，写SQL跑出数来
      - 光看使用人数不行，还要看优惠券下发人数
      - 不同优惠券的消费频次，和优惠券的使用金额
    - 不做提数机器，针对业务单点问题，先追根溯源，建立该类业务的分析框架，由点到面，彻底解决该类问题；同时，在这个过程中，要不断的利用互惠原理和社交技巧，只给业务方做核心的需求，其他的延伸需求让业务方自己动手完成
    - PlanB 找到靠谱的核心的产品经理，了解需求，发现他们的需求时我们的优惠券目前使用情况怎么样，接下来该怎么优化
      - 各类优惠券使用人数
      - 各类优惠券下发人数
      - 各类优惠券使用比例
      - 各类优惠券用户使用时的消费金额
      - 各类优惠券用户的使用频次
- 常规工作优化
  - python日报框架 定期发日报
  - 目标建立
  - 寻找领导帮助

## 4.3 数据异常分析

- 业务理解 指标口径 数据来源都要清楚
- 数据异常排查要形成自己的套路
  - 别人从其它渠道看到的DAU与你提供的不一致
  - 数据口径
    - 移动端 PC端
    - IMEI 手机号
  - 业务理解
    - 数据如何定义 比如DAU
  - 数据产生的过程
    - 数据从哪张表里提取出来的
- 具体套路
  - 验真假 是否真的异常
  - 常见原因，概率降序排列，依次核对
  - 找到原因后 后期要验证
- 判断是否异常套路
  - 问清楚数据咋来的，亲自验证
  - 不要只看近期数据，要看相对长期数据（一个季度）
  - 某指标异常，要看关联指标是否有异常
  - 找人问
- 指标异常常见原因
  - 热点事件
  - 活动影响
  - 政策影响
  - 放假
  - 内部系统问题: 新发版有bug，数据处理过程，传输过程问题
    - 查表中数据量大小
  - 统计口径：业务变更，指标计算方式变化
- 原因确定后的后期验证
  - 持续跟踪数据情况
  - 要有记录，完善文档
  - 发邮件：整个过程分析清楚了再发邮件，描述主要影响和结论
- **异常排查案例** 日活突然暴涨

## 4.4 专题分析

- 举例活动专题分析
  - 需求解读
    - 大于20%的时间分配在沟通需求上，一定要当面沟通
    - 举例，原始需求，负责活动的几个事业群同学，希望看到活动的效果数据
      - 了解需求：属于拉活活动，目前活动对日活的帮助，以及出现的问题有哪些

- 本质需求，活动的拉新促活怎么样，活动带来的用户粘性是否高，哪些活动做得好，哪些活动做得不好，如何优化
- 建立分析逻辑框架
  - 活动运营分析
    - 活动前
      - 流量 活动前每天流量效果，流量波动原因拆解
      - 收入
    - 活动中
      - 流量 DAU(Daily Active User)日活跃用户数量\MAU（即monthly active users）提升
        - 新用户进入
        - 不活跃老用户唤醒
        - 活动期间连续访问情况 用户粘性
        - 各类活动效果对比
      - 收入
    - 活动后
      - 流量 新用户的沉淀  
唤醒老用户后续活跃量  
客服数据，活动复盘
      - 收入
- SQL提数及分析
  - 提数
  - 分析
    - 组成部分 （流量组成部分）
    - 数量比较
    - 有何变化
    - 各项分布
    - 各项相关性
    - 其他深层次挖掘
- 撰写报告
  - 图 文并茂 大部分是图（90%） 少部分是文
  - 先给结论 标题比较重要
  - 阐述的逻辑性要强
- 如何做好专题分析
  - 专题分析要抓住要点，KPI(Key Performance Indicator,关键绩效指标)
  - 周期 2~3周输出完整报告
  - 报告内容通俗易懂，能落地，可执行性强
- 案例：新用户留存专项
  - 背景描述
    - 某产品用户次留，7留与竞品均低5%
    - 新用户流失速度高于竞品

- 希望通过数据分析，找到原因优化产品运营，提升用户留存
- ① 用户留存整体分析
  - 摸清基本数据情况，看看哪些数据有问题，找到优化点
  - 关键点，不要拘泥于细节，出活快



- ② 找到优化点 (1~2个)
  - 关键操作路径数据发现曝光PV到点击PV的CTR（点击率）很低
    - 应该给新用户看什么，什么时候给，放在什么位置
  - 某个量大的二级渠道次日留存率比其它渠道明显偏低
    - 找原因，渠道质量问题，用户是否安装竞品
    - 产品与渠道是否不匹配，对比高留存渠道找差距
  - 给出产品运营建议后，配合AB测试看效果（小部分用户先做实验到最后上线）
- ③ 不断重复前面两个阶段，寻找其他切入点
  - 竞品分析，营销活动分析，用户流失分析
  - 换个角度，跳出当前项目
  - 输出的报告一定得有能落地的东西

## 4.5 小公司日常工作解析

- 靠谱的数据团队：
  - 数仓组
  - 数据产品组（BI开发，日常支持）
  - 数据建模组（模型开发）
  - 专题分析组（临时取数需求，专题报告）
  - 数据管理组（数据口径，数据埋点）

- 实际工作中
  - 横向上，埋点，数据口径，体系指标建立，报表统计，数据清洗，平台研发，专题报告，数据培训
  - 纵向上，对接产品，运营，市场，财务，销售
  - 研发埋点：字段格式，字段名称，字段取值，打点日志，文档沉淀
  - 口径和指标体系梳理：DAU、MAU的定义和计算方法，产品的核心指标及具体业务性指标
  - 平台研发：开发内部大数据平台，数仓，参数开发或者参与数据清洗

## 4.6 数据分析整体流程

- 干活有套路，不能只提数
  - 明确问题：知道问题的上下文
  - 搭建框架：思考问题套路，一 二 三。。。。
  - 数据提取：mysql hive为主
  - 数据处理：Excel R Python为主
  - 数据分析：数据分析的常用套路
  - 数据展示：Tableau excel R Python
  - 撰写报告：文笔功底，整体逻辑性
  - 报告演讲：沟通，表达，报告之前有问题需要提前沟通
  - 报告落地：最有含金量的部分