

# 商业分析核心笔记：从数据洞察到商业行动

## 1. 商业分析的演进与价值链

商业分析并非单一维度的技术，而是一个随着成熟度提升而不断深化的过程。根据

Gartner 的模型，分析能力可分为四个阶段：

(1) 描述性分析 (Descriptive): 发生了什么? (What happened?) —— 基于历史数据生成报表。

(2) 诊断性分析 (Diagnostic): 为什么发生? (Why did it happen?) —— 探究根本原因。

(3) 预测性分析 (Predictive): 将要发生什么? (What will happen?) —— 预测未来趋势。

(4) 规范性分析 (Prescriptive): 我们该怎么做? (What should we do?) —— 提供最优行动建议。

核心价值：拥有高级分析能力的企业，其决策速度比竞争对手快 5 倍。

## 2. 数据基石：颗粒度与数据准备

### (1) 数据的颗粒度

数据的深度决定了洞察的精度。颗粒度指的是数据的细节水平。

- 细颗粒度：能够发现微观趋势和个体行为模式。
- 数据类型：结合分类数据 (Categorical) 与 数值数据 (Numerical) 能丰富可视化效果。
- 定性 vs 定量：定量提供客观基准，定性提供叙事深度。

### (2) 数据清洗利器：Power Query

数据准备通常占据分析师 60%-80% 的时间。Power Query 是 Excel 和 Power BI 中的核心 ETL 工具。

- 核心功能：连接源 (Connect) -> 转换 (Transform) -> 加载 (Load)。
- 关键操作：去重、处理缺失值、数据类型转换、逆透视 (Unpivot)。
- 自动化：通过记录步骤 (M 函数) 实现工作流自动化，无需重复手动清洗。

### 3. 数据可视化与叙事

可视化不仅是绘图，更是沟通。人类处理视觉信息的速度比文本快 60000 倍。

#### (1) 核心原则

- 准确性 (Accuracy): 真实呈现数据，不扭曲。
- 相关性 (Relevance): 以受众需求为导向。
- 清晰性 (Clarity): 去除无关的“图表垃圾”。

#### (2) 高级图表应用

- 箱线图 (Box Plot) & 小提琴图 (Violin Plot): 用于检测异常值 (Outliers) 和查看数据分布密度。
- 散点图与热力图: 发现变量间的相关性。

### 4. 统计分析与预测模型

#### (1) 描述性统计

理解数据的中心趋势 (均值、中位数) 和 离散度 (方差、标准差) 是分析的基础。

偏误 (Bias) vs 噪音 (Noise): 偏误是系统性错误，噪音是随机波动。

#### (2) 预测分析

预测不仅依靠直觉，更依赖模型。

- 时间序列分析 (Time Series): 如 ARIMA 模型，处理趋势和季节性。

- 回归分析 (Regression): 量化变量间的因果关系。
- 工具: SmartPLS 是进行 PLS-SEM (偏最小二乘结构方程模型) 的主流工具, 特别适用于处理复杂的潜在变量关系。验证指标: 关注  $R^2$  (解释力)、P-value (显著性) 和 VIF (共线性)。

## 5. 高级分析: 模拟与大数据

### (1) 模拟分析

当现实世界过于复杂或昂贵而无法直接测试时, 使用模拟。

- 蒙特卡洛模拟 (Monte Carlo): 通过成千上万次随机试验来计算结果的概率分布, 量化风险。
- 离散事件模拟 (Discrete Event): 用于优化排队论、流程瓶颈 (如医院、供应链)。

### (2) 大数据与机器学习

大数据具有 5V 特征: 大量 (Volume)、高速 (Velocity)、多样 (Variety)、真实 (Veracity)、价值 (Value)。

机器学习分类:

- 监督学习 (Supervised): 分类 (Classification)、回归 (Regression) —— 有标签数据。
- 无监督学习 (Unsupervised): 聚类 (Clustering) —— 发现数据内在结构 (如客户分群)。
- 强化学习 (Reinforcement): Agent 在环境中通过奖励机制学习策略。

### (3) 社会网络分析 (SNA)

用于分析节点（人/组织）之间的关系结构。关键指标包括度中心性 (Degree)、中介中心性 (Betweenness) 等，常用于组织架构优化或社交媒体影响力分析。

## 6. 规范性分析：决策的终局

这是分析的最高阶段，回答“我们该如何做”。

核心技术：优化 (Optimization)。

整数规划：用于不可分割的资源（如人员排班）。

连续优化：用于定价策略、投资组合比例。

应用场景：库存优化、个性化推荐、风险评估。