

□ 数据专员培训 & 数据主题课程

知识点汇总与模拟考题

2026年6月

数据专员培训 & 数据主题课程 — 知识点汇总

第一部分：组织架构与业务运营

1.1 新东方集团概况

- **成立时间**：1993年11月16日
- **业务板块**（9大板块）：
 - 素质教育板块：儿童素养智学中心、中学智慧学习中心、高中项目中心
 - 国际教育板块：新东方国际教育、前途出国、斯芬克艺术教育
 - 大学生教育板块：大学生学习与发展中心
 - 教育服务板块：英语学习、教师发展研究院、家庭教育等
 - 智慧教育板块：新东方智慧教育
 - 图书文创板块：大愚文化、私库云书
 - 基础教育板块：满天星、国际双语学校
 - 直播电商平台板块：东方甄选、新东方直播间
 - 文旅项目：新东方文旅
- **规模**（截至2023年2月）：78个城市、748家学校及学习中心
- **太原学校**：2005年成立，1200+员工，三大部门（高中班级部、高中1对1部、青少部），39个教学点

1.2 学校组织架构

- **前台（业务发展）**：青少部、高中班级、高中个性化、文旅研学、其他业务
- **中台（内外客户）**：课程服务部（客服组/教务组）、市场营销部
- **后台（组织保障）**：综合行政部（资产/信息/校办）、财务管理部、人力资源部

1.3 客户旅程

接收宣传 → 课前规划 → 课中课后 → 阶段结课

- **市场宣传**（收资源）→ **学业规划师**（促转化）→ **学习管理师 + 授课老师**（保效果）→ **转介绍/再报名**（传口碑）

第二部分：财务与经营指标

2.1 利润指标维度

- 培训收入 = 班课收入 + 一对一收入
- 班课收入拆解：结转收入 = 在读人头 × 生均小时 × 结转单价
- 一对一收入拆解：结转收入 = 现金流收入 × 结转率
- 成本结构：
 - 主营业务成本：课酬（单小时课酬×班量×单班小时+续班奖金）、教材资料（生均成本×人次）、教室租金
 - 间接费用：行政后勤费、市场推广费、日常运营费、折旧装修费
- 利润 = 培训收入 - 人力成本 - 主营业务成本 - 间接费用
- 利润率 = 利润额 ÷ 培训收入

2.2 太原学校经营数据（5年招生趋势）

财年	暑假入口	暑假非入口	暑假合计	秋季	寒假	春季	总人次	增长率
FY22	42,200	30,470	78,728	33,228	34,083	21,659	158,937	-
FY23	19,574	14,238	37,732	26,198	28,046	23,780	108,353	-32%
FY24	44,127	18,507	67,311	36,508	38,209	34,142	167,042	54%
FY25	59,448	24,968	89,740	50,489	52,693	44,259	227,162	36%
FY26	64,745	35,108	106,149	55,995	59,981	46,372	258,980	14%

- 双减影响：FY23总人次同比下降32%，暑假入口人次下降54%
- 业务复苏：FY24总人次同比增长54%，暑假入口人次增长125%
- 稳健增长：FY25增长36%，FY26增长14%，增速回归稳健

第三部分：校区资产与人力数据

3.1 校区面积管理

- 双减阶段（FY22）：校区面积从6.2万平减至3.1万平（-50%）

- **结构性复苏 (FY23-FY26) :**

- FY24-FY25: 校区面积复合增速41%, 收入复合增速39%
- 校区投入与业绩增长高度同频

3.2 编制管理

- FY24: 收入增速46% VS 编制增速51%
- FY25: 收入增速31% VS 编制增速23%
- FY26: 收入进入稳健增长期, 严格执行冗余资源管控

第四部分：数据标化

4.1 ERP系统

- **班课系统:** 固定时间、多人上课、高性价比 → 班级编码、听课证号、学员编码
- **高端系统:** 时间/科目/教师定制 → 课时包编码、合同编码、学员编码
- **会员卡系统:** 自主学习、线上学习、答疑服务 → 会员卡号、学习小组、会员子卡、学员编码
- **课时包系统:** 流程简单、培养爱好 → 课时包、班级编码、听课证号、学员编码

4.2 标化数据集

班级拼接表

- **业务模式:** 按产品设置规则区分业务条线 (素养、纯学习机、学习机托管等)
- **财务模式:** 与集团财务命名保持一致
- **核心字段:** 班级编码、班级名称、班级状态、产品体系、管理项目、标准部门
- **标签字段:**
 - 年级序数处理值: 统一命名规则, 多值记"不区分"
 - 科目处理值: 统一科目命名 (如"编程: 机器人、编程")
 - 合班标签: 完全合班/其他合班/未合班
 - 合并上下册班级编码: 将春秋上下册合并
 - 业务季度: 20241(暑假)/20242(秋季)/20243(寒假)/20244(春季)
 - 长段标签: 预计时长较长为长段班
 - 带课最多教师

- 开课/结课日期处理值
- 班容分类：1对1/6-8人/12-18人/20-30人/35-45人/50人以上
- 开课时点人数/财季时点人数/结课时点人数
- 授课方式标签：常规面授/在线直播/在线录播
- 教学区处理

花名册拼接明细表

- 进班/离班自然月、财月、财季、财年
- 开课流失标志（实际缴费课次 ≤ 0 ，且同季度同科目无有效在班记录）
- 开课流失标志（实际缴费课次 > 0 ，且同季度同科目无有效在班记录）
- 开课退费/开课后退费

在读拼接表

- 结转开始月、结转结束月
- 学员在读日期对应的自然月/自然年/财月/财季/财年

收入拼接表

- 收入类型：现金收入、预计结转收入、结转收入
- 自然月/自然年/财月/财季/财年

留存续费扩科宽表

- 多季联报最晚结课班级编码
- 原班联报班级最晚结课日期/自然月/财月/财季/财年

第五部分：收入指标

5.1 现金收入

- **定义**：某订单内产生的全部费用（含所有关联班级）
- **公式**：现金收入 = 收入金额 - 退费金额（按订单汇总）
- **时间维度**：下单时间、首个开课班级开课日期、下单财季
- **场景**：顾问端的签单收入

5.2 开课收入

- **定义**：某个班级内产生的全部费用
- **公式**：开课收入 = 现金收入A = 收入金额A - 退费金额A（按班级汇总）
- **时间维度**：下单时间、班级开课日期、下单财季、班级归属业务季度
- **场景**：教学端每个班级的招生收入

5.3 预计结转收入

- **定义**：班级在行课财周/月/季内产生的税前或税后结转收入，包含未实际结转的收入
- **计算**：按财务规定的预计结转计算规则汇总求和

5.4 结转收入

- **定义**：班级在行课财周/月/季内实际产生的税前或税后结转收入
- **计算**：与预计结转收入计算规则相同，但仅计算已过时间周期内的收入

5.5 实际缴费金额

- **公式**：应报名收入 - 实际优惠金额（退费则计为0）

5.6 实际消费金额

- **公式**：实际缴费金额 - 实际退费金额

5.7 实收小时单价

- **公式**：实际消费金额 ÷ [(班级预计时长 ÷ 60) × (实际上课课次 ÷ 班级课次)]

第六部分：招生指标

6.1 核心算法

- **班级维度人次**：非一对一按听课证号去重，一对一按学员编码去重
- **汇总维度人次**：非一对一按(学员编码+科目)去重，一对一按学员编码去重
- **人数**：按学员编码去重计数
- **数据范围**：仅常规体系班级，会员卡不含体验卡，一对一不含定金/升年级/赠送

6.2 报名人次

- **定义**：以是否产生报名动作判断学员是否有效

- **时间维度：**进班日期、进班财月、进班财季

6.3 开课人次

- **定义：**某业务季度内开课班级的当前有效报名人次（在班状态为有效）
- **时间维度：**业务季度、业务归属年

6.4 招生人次

- **定义：**某季度某科目下当前有效报名人次（按汇总维度）
- **时间维度：**业务季度、业务归属年

6.5 在读人次

- **定义：**某段日期范围内有实际行课的学员
- **判断方法：**行课开始=学员进班日期与实际开课日期的最晚值；行课结束=离班日期与实际结课日期的最早值

6.6 新生/老生人次

- **新生：**首次报名或跨业务季度报名某科目（生源类型≠留存老生）
- **老生：**连续业务季度报名某科目（生源类型=留存老生）

6.7 招生人次周进量

- **公式：**本周进班人次 - 本周离班人次

第七部分：留存指标

7.1 开课前流失

- **条件：**退费学员，同业务季度同科目无其他常规体系有效记录，且实际缴费课次 ≤ 0
- **开课前流失率** = 开课前流失人次 \div (开课前流失+开课后流失+招生人次)

7.2 开课后流失

- **条件：**退费学员，同业务季度同科目无其他有效记录，且实际缴费课次 > 0
- **开课后流失率** = 开课后流失人次 \div (开课后流失人次 + 招生人次)

7.3 留存率

- **原班人次**：行课且实际缴费课次>0的学员，按科目+学员编码去重
- **留存人次**：原班学员中，续报下一连续季度同科目班级的学员
- **留存率** = 留存人次 ÷ 原班人次

7.4 续费率

- **应续费人次**：本季度班级结课学员（单季度）或联报中最晚结课班级所在业务季度的学员（联报）
- **已续费人次**：续读到下一连续业务季度同科目班级/会员卡
- **续费率** = 续费人次 ÷ 应续费人次
- **多季联报**：同一订单报名多个连续季度班级，或同一天报名，或首班开课前报名多个季度

第八部分：产能指标

8.1 实际开班量

- **规则**：不含0人班/取消/未审核；完全合班算1个；子母仅统计母班；上下册仅统计长段

8.2 预招人次

- 未超售按标准人数，已超售按当前人数

8.3 班均人数

- 开课人次 ÷ 实开班级数量

8.4 满班率

- 开课人次 ÷ 预招人次

8.5 教师产能

- **带课教师人数**：排过课的教师，不含当季入职/离职/三期教师
- **教师带课量**：配课表课时按财季加总（不含录播课，合班算一次）
- **师均带课量** = 教师带课量 ÷ 带课教师人数
- **教师带课量中位数**：按带课量排序取中间值

8.6 教室产能

- **教室课量**：可用教室已承载/预计承载课量（不含线上课）

- **教室可用小时数**：正常使用教室，平时2小时/节假日10小时
- **教室使用率** = 教室课量 ÷ 教室可用小时数

第九部分：数据产品

9.1 大屏产品

- **实时续费大屏** → 会员卡续费大屏
- **招新大屏** → 招新数据大屏

9.2 管理报表

- **校长运营管理驾驶舱**
- **BR集团标准化报表** → 业绩进展、驾驶舱

第十部分：数据分析方法

10.1 数据分析步骤

- **确定分析目的**：指标与维度（时间、新老用户、年级、科目、渠道）
- **数据获取**
- **数据分析**：对比细分、生命周期分析、漏斗分析、相关性分析、5W2H、逻辑树、RFM
- **结果可视化**：“文不如表，表不如图”
- **总结与建议**

10.2 对比细分

- **多维度**（时间/用户/地区/构成）
- **与上周、上季度、去年同期对比**
- **保证单一变量原则**

10.3 生命周期分析法

- **分位数法**：首次退费到再次报名间隔天数的90分位值
- **拐点法**：观察留存在第几周趋于稳定

10.4 RFM用户分群法

- **R (Recency)** : 最后一次付费距今天数 → 消费习惯
- **F (Frequency)** : 付费频次 → 消费意愿
- **M (Monetary)** : 付费金额 → 消费能力
 - 分群矩阵: 重要价值/重要召回/重要发展/重要挽留/潜力/一般维护/新用户/流失

10.5 相关性分析

- 正相关: $|R| > 0.95$ 显著, $|R| \geq 0.8$ 高度, $0.5 \leq |R| < 0.8$ 中度, $0.3 \leq |R| < 0.5$ 低度, $|R| < 0.3$ 极弱
- 负相关: $R < 0$
- 无线性关系: $R = 0$

10.6 麦肯锡逻辑树分析法

- **拆分原则**: 要素化 (归纳相同问题)、框架化 (不重不漏MECE)、关联化 (保持相互关系)
- **示例**: 提升学员规模 = 增加招新 + 降低流失

10.7 漏斗分析法

- 直观呈现用户行为各步骤之间的转化率

模拟考题 (10题)

题目1: 招生人次计算 - 暑假班级

背景: 2024财年暑假 (业务季度20241), 英语科目班级A的报名情况:

- FY23Q4有2名学员A和B以正价报入
- FY23Q4有1名学员C以定金形式报入计费体系班级, FY24Q1正式转入正价班级A
- FY24Q1有3名学员E、F、G以正价报入, 但学员G在开课后退班

问题:

1. 分别计算班级A在FY23Q4和FY24Q1的报名人次

2. 计算班级A在FY24Q1的开课人次

提示：报名人次以产生报名动作为准；开课人次需在班状态为有效。

题目2：在读人次计算

背景：3月10日至6月10日行课的春季班级，共有A、B、C、D、E、F五名学员报名。学员A开课前提退班，学员E行课至5月30日退班，学员F行课至6月5日退班。

问题：

1. 计算该班级在Q4财季（3月-5月）和Q1财季（6月-8月）内的在读人次
2. 计算该班级春季业务季度内的在读人次

提示：行课开始=学员进班日期与实际开课日期的最晚值；行课结束=离班日期与实际结课日期的最早值。在读：在该时间段内有行课记录的学员。

题目3：收入计算

背景：2024财年暑假业务季度内某班级，共有A、B、C、D、E五位学员报名，每位学员实际报名费用1000元。

- 学员A和B在FY23Q4报名
- 学员B在FY24Q1全额退款
- 其余学员在FY24Q1报名且未退费

问题：

1. 求该班级在FY23Q4和FY24Q1的现金收入
2. 求该暑假班级的开课收入

提示：现金收入按订单汇总（收入金额-退费金额）；开课收入按班级汇总。

题目4：实收小时单价计算

背景：某班级预计时长720分钟共8课次：

- 学员A：开课前提前报名，全程未退费，实际消费金额800元
- 学员B：第3次课插班报名，报名后未退费，实际消费金额500元

- 学员C：开课前提名，行课至剩余2课时时退费，实际消费金额600元

问题：计算每位学员的实收小时单价。

提示：实收小时单价 = 实际消费金额 ÷ [(班级预计时长÷60) × (实际上课课次÷班级课次)]

题目5：续费率计算

背景：某学校2024财年暑假业务季度（20241）结课后，有500名应续费学员。其中：

- 300名学员续费到20242秋季同科目班级
- 50名学员续费到不同科目班级
- 其余学员未续费

问题：

1. 计算该暑假季度的续费率
2. 如果多季联报的学员中有80人原本是联报（已在联报时完成续费），他们是否计入续费人次？为什么？
3. 假设上述80人联报学员也算作续费（即已续费人次变为380），续费率变为多少？

题目6：留存率与流失率计算

背景：某业务季度内，某科目共有1000名学员报名。数据如下：

- 班级实际缴费课次>0的学员：800人
- 续报下一季度同科目班级的学员：600人
- 退费学员中，实际缴费课次≤0且无其他有效记录：50人
- 退费学员中，实际缴费课次>0且无其他有效记录：80人

问题：

1. 计算留存率
2. 计算开课前提失率
3. 计算开课后流失率
4. 开课前提失+开课后流失+招生人次是否等于原班人次？请验证。

题目7：招生人次周进量分析

背景：某学校连续4周的招生数据如下：

周次	本周进班人次	本周离班人次
第1周	200	30
第2周	180	45
第3周	220	35
第4周	150	60

问题：

1. 计算每周的招生人次周进量
2. 第4周相对于第1周，周进量下降了多少？
3. 如果你是数据分析师，面对第3周到第4周的趋势，你会从哪些维度做对比细分分析？

题目8：班均与满班率计算

背景：某学校20241财季（暑假）招生数据：

- 实开班级数量：50个
- 开课人次：1,200人
- 预招人次（按标准人数计算）：1,500人
- 实际超售班级：有5个班级超售至各自的班级最大人数（超售总额外人数80人）

问题：

1. 计算班均人数
2. 计算满班率
3. 如果考虑超售因素，重新计算满班率（假设标准预招人次1,500中未包含超售人数）

题目9：相关性分析与逻辑树

背景：某校区想提升学员留存率，收集了80名学员的数据：

- 学员出勤率与留存率的相关系数 $R = 0.82$

- 学员NPS评分与留存率的相关系数 $R = 0.45$
- 退费天数与留存率的相关系数 $R = -0.35$

问题：

1. 分别说明上述三个变量与留存率的相关性强度和方向
2. 使用逻辑树分析法，分解"提升学员留存率"这个目标，至少列出3个二级拆分维度
3. 如果出勤率和留存率高度相关，你建议校区优先从哪个方向入手？

题目10：综合数据分析 - 校区经营诊断

背景：太原学校2026财年暑假（20241）的关键数据如下：

指标	数据
暑假入口人次	64,745
暑假非入口人次	35,108
暑假总人次	106,149
同比FY25暑假增长	18.3%
实开班级数	2,600个
带课教师人数	320人
教师总带课量（课时）	38,400课时
教室可用小时数	52,000小时
教室课量（已承载）	41,600课时
FY24暑假班均人数	44.3人/班
FY25暑假班均人数	42.5人/班
FY26暑假满班率（按标准1,500人）	需计算

问题：

1. 计算FY26暑假的班均人数
2. 计算师均带课量
3. 计算教室使用率
4. 对比FY24（班均44.3）、FY25（班均42.5）和FY26的班均趋势，说明该数据说明什么？

5. 如果下季度预算编制要求利润提升10%，利用逻辑树分析法，列出三条可能的提升路径（每条路径需包含具体指标）。
