



## 2020 年百威校园挑战赛常见问题

### 一. 关于赛事规则和组织

问题 1: 关于一轮任务中的三分钟视频, 请问需要参赛队伍选手露脸吗? 是否可以直接以 PPT 为画面进行相关思路的阐述呢?

回答: 视频内容为团队对于所提交 ppt 进行的讲解和阐述; 视频要求团队三人均出镜讲解, 但不要求必须在一个镜头出现; 当然到底是在三分钟内选手是一直出镜, 还是只是在前边、中间或者后边出镜由队伍自己决定。我们的考虑因素是这是一个团队任务, 我们希望在考察提交作品质量的同时, 也有机会考察团队在比赛中的分工、协作、配合和综合实力。

问题 2: PPT 5-10 页 包含封面和尾页吗?

回答: PPT 总页数不超过 10 页, 包含封面和尾页。

### 二. 关于数据建模方向

问题 1: Price 三分之一的缺失值是官方有意为之? 还是单纯没有这三分之一的数据?

回答: 部分数据由于业务原因可能存在缺失, 需要选手去判断是否可能修补或是否需要使用这个变量的数据。

问题 2: 评委对可视化 tool 是否有偏向?

回答: 无, 欢迎使用任何可视化和分析的工具。

问题 3: 预测结果要求是一个准确值还是一个范围?

回答: 预测结果要求是一个准确值, 而不是一个范围。

问题 4: 一定要用 PPT\Excel 做可视化吗?

回答: 对选手做可视化和建模的工具没有任何限制。可视化和分析的结果截图贴到 PPT 用于提交作品。

问题 5: 第一轮比赛要出预测结果吗?

回答: 第一轮比赛不需要提供预测的结果。

问题 6: 可以使用外部数据建模吗?

回答: 非常鼓励选手获取外部数据进行分析与建模。同时也需要选手判断我们提供的数据是否对销量的预测有帮助。

问题 7: 粗略建模是什么意思?

回答：非常鼓励选手获取外部数据进行分析与建模。同时也需要选手判断我们提供的数据是否对销量的预测有帮助。

问题 8：对第一轮视频和 PPT 的预期是什么？

回答：可视化图表加一定的分析说明。

问题 9：首轮不需要给出具体数据？

回答：不需要出具体的预测数据。

问题 10：Low-end2 和 3 是并列关系还是递进？

回答：并列关系，只是一个序号。

问题 11：模型结果在评分中占比多少？有一个标准评分的公式吗？

回答：预测结果的评估标准将在第二轮主题发布时给出。

问题 12：10 页 ppt 对每页字号和图表数有要求吗？

回答：无。

问题 13：对特征直接重要性和相关性的打分点怎么样。

回答：分析的角度内容，各选手可根据自己的项目。

问题 14：数据里 Alu bottle 和 big bottle 哪个容量更大？

回答：小听正常为 330 毫升，小瓶 200-330 毫升，大听 500 毫升，大瓶约 500-600 毫升，铝瓶 350 毫升。

问题 15：第一轮需要给预测值吗？

回答：不需要。

问题 16：比较倾向中文还是英文？

回答：语言不作要求，重点是思路和作品。

问题 17：视频需要讲解 ppt 吗？

回答：视频内容可为录屏对数据可视化分析成果的讲解和建模思路的阐述。

问题 18：销量为什么会是 10 点几啊？

回答：单位为百升。

问题 19：PPT 模版和动画会影响评分吗？

回答：不会。

问题 20：评分是对每一轮叠加吗？

回答：没有。

问题 21：有些产品只在某一些地区有销量数据，是表示在另一地区销量为 0 吗？

回答：是的，并非所有产品在所有地区都有销售。

问题 22：需要预测每个产品在每个地区的销量吗？

回答：是的，会在第二轮给选手提供需要提交预测的区域和产品模板。

问题 23：可以建立两个模型对应两个地区吗？

回答：可以。

问题 24：关于 Commercial Planning 中的字段说明：

回答：若 YM 为 202001，即为在 2019 年 12 月可获得的 2020 年 1 月的商务费用投入预估。可以简单理解如下字段都是为了销量服务的投入，不需要太纠结单位和深层的含义，以关注变量和目标值之间的关系为主。

字段	解释	例子
YM	投入月份	202001
SalesRegion	销售区域名称	Heilongjiang
SKU Code	产品代码	1, 2, 3, 4
Price	每箱单价/人民币	909
Rebate	预估返利金额投入/人民币	3882833
PromotionVolume	预估促销量/百升	24388
Discount	预估折扣投入/人民币	187893947

问题 25：关于预测业务场景和逻辑的说明。

回答：即在 M0，使用 M-1 月截止的历史销量，和在 M0 可获得的对 M+1 商务费用投入的预估和任何可取得的外部数据，来预测 M+1 的销量。如在 2019 年的 6 月，需要使用截止 2019 年 5 月 31 日的历史销量，和 201907 的商务费用投入计划，来对 2019 年 7 月的销量进行预估。在获取数据，做特征工程时需要关注这个逻辑。

### 三. 关于商业方向

问题 1：在所给数据的“菜品数据”一表中，有啤酒品牌、啤酒生产商两栏，我们对分析对象存有一些疑问：请问是需要以“啤酒品牌”中的“BUD”作为主体分析，还是以“啤酒生产商”中的 ABI 为对象作为主体分析呢？因为“门店数据”中特别给出了 ABI 啤酒的销售数据，因而我们有此疑问，希望您能慷慨解答，谢谢！

回答：在我们的任务中只提及了一种表述，即 ABI 啤酒，BUD 是 ABI 旗下品牌之一。

问题 2：关于 commercial 赛道比赛第一轮任务里面提出的维度中，对铺店率这个名词不理解，在网络上也找不到相关名词的解释，可以理解为连锁店的数量吗？

回答：从“率”的表述可知此概念涉及分子和分母，铺店率我们关注 ABI 啤酒铺开到店铺售卖的程度。

问题 3：commercial guide 中提到“提升模拟门店百威亚太啤酒销量的初步思路”，可以明确一下模拟门店的定义吗，是在外卖平台上的店铺（一般不是专门卖啤酒的），还是像百威的小程序上买啤酒即时达那样的门店呢？

回答：当第一轮我们分析完行业概况后，会在第二轮给到若干家外卖门店数据信息进行模拟经营。

问题 4：门店数据 F 列店铺评分为 0 的是没评分还是评分为 0？菜品数据表单中的 G 列商品价格是啤酒单品价格，还是 F 列商品分类的一些套餐价格？

回答：这是大数爬取的结果，0 一般为无数据。F 列商品分类仅指该商品被分在了什么类目，有的时候啤酒单品会被分到“套餐”的类目，但 G 列的价格就是啤酒单品价格。从“率”的表述可知此概念涉及分子和分母，铺店率我们关注 ABI 啤酒铺开到店铺售卖的程度。

问题 5：模拟门店是指可以有我们决定菜品地理位置的线下门店还是指线上的电商门店？具体的评分或晋级标准？

回答：模拟门店信息来自我们给定的若干家线下外卖门店。会根据完整度、创意度、专业度综合考虑后打分从“率”的表述可知此概念涉及分子和分母，铺店率我们关注 ABI 啤酒铺开到店销售卖的程度。

问题 6：请问在“门店数据”中的菜系的划分是默认不重叠的吗？比如“川湘菜”是否包含“香锅干锅”，“地方菜系”是否包含“川湘菜”呢？谢谢！

回答：菜系的划分是不重叠的，一个门店被定义为“川湘菜”就不是“香锅干锅”；同理，“地方菜系”不包括“川湘菜”，“地方菜系”与“川湘菜”分设即代表“地方菜系指非川湘的其他菜系”。

问题 7：请问“在这一轮，每支队伍与需提交 1)5-10 页的数据分析/数据可视化 PowerPoint 报告，您对百威亚太啤酒餐饮外卖生意现状的洞察，以及提升模拟门店百威亚太啤酒销量的初步思路”中的“啤酒销量”，指的是啤酒销售的支数，还是啤酒销售的体积（HL），还是啤酒销售的金额呢？谢谢！

回答：我们的最终目的是寻求卖出更多酒的战略，所以在你所提及的单位中请选择一个最能表示卖出酒的量的。

问题 8：美团和饿了么上的“果蔬商超”类别的外卖业务，能不能算是餐饮类的外卖业务？

回答：不计算在内，餐饮类在我们的表述中为 CR, chinese restaurant。

问题 9：原始数据中 price\_num 的缺失值很多的原因是什么，该项缺失值是否有特殊含义？

回答：促销价是一个商品结算的价格，比如一个商品打折了一定在付款时以折后价支付，商品价缺失即表示该商品未促销。

问题 10：我们尝试提升的，仅仅是“百威啤酒”的销量，还是囊括哈尔滨啤酒等品牌的百威旗下所有啤酒品牌的销量？谢谢！

回答：旗下所有，即 ABI 啤酒销量，百威啤酒称呼为 BUD。

问题 11：商品的促销价格是否为商品的平均成交价格？

回答：促销价是一个商品结算的价格，比如一个商品打折了一定在付款时以折后价支付。

问题 12：数据中存在同一连锁名但连锁店数量不一致的情况，是否为原始数据存在问题？

回答：查看平台列，这两个数据分属饿了么与美团。

问题 13：在菜品数据 Sheet 中的 F 栏（sku\_cls\_desc）中有一些字符后会标注“单点不外送”，那么没有标类似字句的是否可以默认为“单点外送”？

回答：这是直接从商家处爬虫的数据，即商家在该类目标注了“单点不外送”，请结合点外卖的实际生活场景进行判断。

问题 14：请问是否能引用自查数据进行补充？

回答：可以在展示时提及通过自查什么数据，得到了哪些分析结论。

问题 15：关于提升模拟门店百威亚太啤酒销量的问题，这个模拟门店，请问是一家普通的模拟餐饮店，还是百威的一家模拟自营店呢？

回答：模拟门店是我们给定的若干家外卖门店的销售信息，数据来自三方爬虫公司。

问题 16：在进行数据分析的时候，我们发现数据表-门店数据 sheet 中有数据有出入。如 M 列 poc\_chain\_num（连锁店数量）的第 chain42 在第 75 行、377 行等对应的数据是 2778，而在其他行对应的数据则是 3108，请问我们应该以哪一个为准呢？

回答：查看平台列，这两个数据分属饿了么与美团。

问题 17：表格的数据收集有没有特定地区一说？例如这个数据是来自上海的，或者说是北京的，还是说全国的？又或者说您希望我们用 A 地区来代替这个？另外就是我们自己是否可以在私下中做一些问卷调查来扩充客户数据和理由来让我们的报告具有多样性？

回答：我们不给出本数据特定的区域信息，所以如果做问卷调查可能不能锚定某区域。

问题 18：模拟门店是指根据调查和分析寻找最适合的外卖门店场景？或是指别的情况？

回答：模拟门店是我们给定的若干家外卖门店的销售信息，数据来自三方爬虫公司。

问题 19：附件中 EXCEL 的文件数据是否默认为代表亚太地区的情况，我们以此为基本针对亚太地区提出意见？还是我们仅需针对中国地区的情况提出意见？

回答：饿了么与美团两个平台的数据，中国地区。

问题 20：我们参与的是商业方向的挑战，想请问一个关于表格中字段含义的问题。在“门店数据”表中，“poc\_chain\_name\_dummy”以及“poc\_chain\_num”这两个字段对不上号，也就是在数据中会存在这样的现象：一个连锁名会对应几个不同数据的连锁店数量，请问连锁名和连锁店具体是指什么意思呢？谢谢！期待您的回复！

回答：连锁门店名字是连锁门店的名称总称，连锁门店数量是描述在这个连锁名字下面门店的数量。

问题 21：模拟门店：请问模拟门店的经营方向是我们自己选的吗，是我们以模拟门店的商販为角度为出发点提建议，还是我们是百威公司的人去经营这家模拟门店为角度来写商业计划呢？

回答：模拟门店是百威公司提供真实的门店给选手用作虚拟操作，需要从卖啤酒角度提升啤酒销量。

问题 22：请问给出的销售量数据为疫情期间的某个月的月销售量吗。给出的啤酒总销量（HL）也为月销售量吗？

回答：2020 年某个月啤酒销量。

问题 23：在三十个以上的门店数据中，都出现产生月售单数为 0 但是全部产品订单金额数不为零的情况。如图中所示的所有门店以及 poc2571 门店等等。出现此类情况的门店数量预计在五十个左右。此外，我们用全部产品订单金额数除以月售单数计算了每家店平均每单的金額，发现 130 家店的平均每单外卖超过 1000 元，希望可以重新确认此数据的真实性？

回答：反常数据可能是异常值，请剔除。

问题 24：我想问一下这次的方案针对百威亚太旗下的全部啤酒种类，还是针对百威啤酒这一个种类？还是我们选择一种？

回答：针对百威旗下整个啤酒品类。

问题 26：百威模拟门店是只包含线上还是线上线下都包括呢？

回答：只包括线上。

问题 27：【该店产生月售单数（order\_num）】是否不包含在【该店全部产品订单数量加总（sales\_volume\_num）】？

回答：月售单数是门店销量信息不止是啤酒销量包含所有销量，sales volume 是啤酒销量信息。

问题 28：beer\_sales\_by\_volume\_num(hl) 中的 volume 是不是升 跟后面的 volume 销量不是一个意思？

回答：是一个意思。

问题 29：请问具体的评分标准是什么呢？

回答：根据方案优劣评分。

问题 30：想问那个校园要做的是百威这个单个品牌还是百威亚太这个总的品牌？

回答：啤酒品类是分析的重点，覆盖百威旗下所有品牌。

问题 31：请问给的数据是什么数据呢？

回答：真实的脱敏数据，可以代表中国啤酒外卖情况。

问题 32：商业方向，对于数据分析使用模型方向有什么要求吗？

回答：根据各自方案不同准备各自模型，没有强硬要求。

问题 33：数据会有异常出错的值吗？

回答：会出现，请排除异常值。

问题 34：外面平台有别的数据么？

回答：请勿使用其它外部数据。

问题 35：请问百威和百威啤酒是同一种产品吗？

回答：是一种品类。

问题 36：策略必须基于两个数据表内的变量提出吗？

回答：策略必须结合数据学习结果以及消费者洞察得出。

问题 37：请问目标值是指门店销售金额吗？

回答：是啤酒销量。

问题 38：同一家点的订单数是一天内的订单吗？

回答：一个月数值。

问题 39：我们展示是大体思路还是具体方案？

回答：大体思路。

问题 40：表格数据是特定地区的还是整个中国大陆的呢？



回答：大中华区代表性数据。

问题 41：所以我们是给出建议为主，数据分析为辅助吗？还是数据分析为主？

回答：两者不可偏废，既需要数据学习结果，也需要商务洞察分析。

问题 42：评分为 0 是什么意思呢？

回答：可能是异常值。

问题 43：这个门店是外卖平台的吗还是说实体门店有外卖服务呢？

回答：外卖平台门店数据。

问题 44：所以模拟门店是需要我们在表格里选择一家是吗？

回答：不是，是公司会提供一家实际门店来按照参赛者想法更改运营参数。

问题 45：数据中给的啤酒销量包含 ABI 啤酒的销量吗？

回答：包含。

问题 46：啤酒品类也互不包含么？

回答：包含。

问题 47：表格的数据是不是 19 年的或是更早的？

回答：2020 年某个月数据。

问题 48：啤酒订单是买一支算一个订单吗？

回答：订单维度是指实际外卖订单，不一定只有一支啤酒。

问题 49：请问自己做的消费者调研数据可以用吗？

回答：建议做。