**《项目工作记录表》**

*教务老师填写：*

|  |  |  |
| --- | --- | --- |
| 学生姓名：李子骏 | 院校：伯明翰大学 | 专业：计算机科学 |
| 项目企业： | 岗位： | 项目周期： |

一、每周工作记录

*学生填写：*

|  |  |
| --- | --- |
| 第一周 工作记录 | |
| **请详细陈述本周的工作内容：**  Tip: what have done + what have achieved + tools  例： 2021年1月1日～1月7日，本周共完成到2个工作任务： 1.完成新能源汽车行业的市场细分研究： 首先我根据阅读在知网和ResearchGate查到的相关论文材料，以及某新能源汽车品牌的企业年报进行了初步市场调研。之后我用Python Matplotlib对初步调查结果进行了数据可视化，分析了近年来新能源汽车生产和消费的的趋势变化，从而发现201？～201？年间的产量和销量有大幅上升等等。交付了一份15页的行研报告。  2.完成该行业某客户企业的战略分析咨询： 在进行过市场研究的基础上，我以问卷的形式调查了新能源汽车消费者和潜在消费者的心理和行为。结合上一步完成的市场调研，运用了本科期间所学的组织行为学、微观经济学和定量分析，提出了发展战略建议。 | 2023.3.28~2023.4.3,   电商平台用户表现现状的分析:  1. 基础信息：日PV有多少（浏览人次），日UV有多少（浏览去重人数），付费率情况如何（付费人数/浏览人数），复购率是多少（购买2次以上人数/购买人数），漏斗流失情况如何（浏览、收藏、加购物车、购买每一步漏斗的比例）  2. 在哪一天发生了购买量的突增和突降，可能的原因是什么（指标拆解和维度拆解） |
| **本周工作中的一些个人小收获（知识、技能、沟通能力、职业素养等方面）：**  例： 本周通过导师辅导与资料查阅，我了解了咨询的工作性质和内容，比如为行业或企业发展出谋划策，为行业或企业实现目标提供解决方案等。  在信息检索方面，我掌握了利用关键词在中国知网、万方数据等数据库查阅信息、并通过扫读文献摘要筛查高度关联性文献的能力，获取了过去3年整个汽车行业的发展状况、新能源汽车的发展状况、消费者对新能源汽车的态度及观点以及新能源车行业的趋势变化等4个方面的相关资料和数据。  同时，本科期间的课程知识也在实习研究中得到发挥：我运用国际商务专业定性管理研究方法、商业和市场经济原理、管理学和金融、营销等专业知识研究以上四个方面的问题，今后在工作中也会充分发掘利用已有的知识技能储备。 | 了解了对数据的指标拆分和维度拆分, 学习了伯克森悖论与辛普森悖论  在信息分析方面, 我掌握了使用excel建立表格进行数据分析和折线图生成. 同时也学会了如何利用pandas和matplotlib库, 在python当中快速整理和生产相关数据 |

|  |  |
| --- | --- |
| 第二周 工作记录 | |
| **请详细描述本周的工作内容：** | 2023.4.4~2023.4.10, 目前为了评估补贴策略优化效果，希望基于实验数据进行分析（实验组A为优化策略，实验组B为现有策略）：   1. 实验分流是否均匀？实验组A人群的转化率和购买人群人均购买金额是否显著高于对照组？ 2. 哪类人群在优化策略上提升更为显著？ |
| **本周工作中的一些个人小收获（知识、技能、沟通能力、职业素养等方面）：** | 学习了： 对照组和实验组的重要性：实验设计中的对照组和实验组是非常重要的，它们可以帮助我们比较不同条件下的结果，确保我们的实验结果具有可靠性和可重复性。  t检验和z检验的原理：t检验适用于小样本量（样本量小于30）的情况下，可以检验两组样本均值之间是否有显著差异；z检验适用于大样本量（样本量大于30）的情况下，可以检验两组样本均值之间是否有显著差异。  如何运用Python库进行t检验和z检验 |

|  |  |
| --- | --- |
| 第三周 工作记录 | |
| **请详细描述本周的工作内容：** | 2023.4.11~2023.4.17 目前为了评估广告效果，希望基于广告曝光数据进行分析（广告曝光用户数据集）：   1. 在全量数据及（data\_all.csv）中找到与广告曝光用户相似的用户，并看用户在不同特征上的分布差异？ 2. 广告曝光用户人均消费是否高于未被广告曝光且相似用户？在哪个行为类型（behavior\_type）的专户上更高？ |
| **本周工作中的一些个人小收获（知识、技能、沟通能力、职业素养等方面）：** | 学会了PSM（Propensity Score Matching，倾向性匹配）方法是一种用于处理非随机实验的方法，主要用于降低处理组与对照组之间的偏差，通过调整个体的倾向得分，使得处理组和对照组之间更加平衡，减少实验因素和其他因素对结果的干扰。  在使用PSM方法时，需要先构建一个分类器（如logistic回归模型）来预测每个被试进入实验组的概率，得到每个被试的倾向得分。然后根据倾向得分将实验组和对照组进行匹配。匹配后，可以用一系列的统计分析方法来分析实验结果。 |

|  |  |
| --- | --- |
| 第四周 工作记录 | |
| **请详细描述本周的工作内容：** | 2023.4.18 理解常见的数据分析大模型， 并学习完整的数据分析流程 |
| **本周工作中的一些个人小收获（知识、技能、沟通能力、职业素养等方面）：** | 本周学习了uplift模型及其在数据分析中的应用。uplift模型的主要目的是通过比较处理组和对照组的效果，确定推广活动是否真正对目标群体产生积极影响，进而实现精准营销。 |

二、项目工作总结

*项目结束后填写：*

|  |  |
| --- | --- |
| **请对项目期间的总体工作内容进行描述：** | 在项目期间，我主要负责分析电商平台用户的表现现状，评估补贴策略优化效果，以及分析广告效果。在分析用户表现时，我学习了伯克森悖论和辛普森悖论，掌握了使用Excel和Python进行数据分析。接下来，在评估补贴策略优化效果时，我了解了对照组和实验组的重要性，学会了运用Python库进行t检验和z检验。最后，为了评估广告效果，我学会了倾向性匹配（PSM）方法，以减少实验因素和其他因素对结果的干扰。 |
| **项目期间工作中的个人小收获（知识、技能、沟通能力、职业素养等方面）：** | 项目期间，我在知识和技能方面有了很多收获。首先，我学会了如何在数据分析中使用Excel和Python，熟练掌握了pandas和matplotlib库。其次，我学会了进行实验设计时的对照组和实验组的设置，以及如何使用t检验和z检验进行数据分析。此外，在广告效果评估方面，我学会了使用倾向性匹配方法。 |
| **请说一说你在项目工作中遇到了哪些困难？**  **以及通过哪些具体的方式、方法，解决了这些问题？** | 在项目过程中，我遇到了一些困难。首先，在进行数据分析时，我发现数据量较大，使用Excel处理起来效率较低。为了解决这个问题，我学习了Python中的pandas库，使用它进行数据处理，提高了分析效率。其次，在评估广告效果时，我发现实验组和对照组之间存在较大偏差，影响了实验结果的准确性。为了解决这个问题，我学习了倾向性匹配方法，通过调整倾向得分使实验组和对照组之间更加平衡。最后，在分析补贴策略优化效果时，我遇到了如何判断实验组与对照组之间差异显著性的问题。我学习了t检验和z检验的原理，并使用Python库进行了相应的检验，解决了这个问题。 |