

黄湘云

求职意向：数据分析师、商业分析、数据科学、数据研发岗位

曾就职于美团搜索技术部、51Talk 数据智能中心，从事数据分析、数据挖掘和报表开发工作。熟练使用查询语言 SQL，协作工具 Git 和统计分析编程语言 R，了解 Spark、Hive 等大数据开发工具。多届中国 R 语言大会演讲嘉宾，曾担任统计之都副主编，软件工具栏目编辑。与赵鹏、谢益辉出版专著《现代统计图形》。

教育经历

- 2011.09
|
2015.06
- **中国矿业大学 (北京)**
理学学士，理学院数学与应用数学专业 📍 北京, 中国

论文: 统计方法在区域经济社会发展综合评价中的应用

 - **中国矿业大学 (北京)**
理学硕士，理学院统计学专业，数据分析和统计计算方向 📍 北京, 中国

论文: 空间广义线性混合效应模型及其应用

工作经历

- 2022.05
|
现在
- **自由职业**
写作、读书、沉淀 📍 北京, 中国
 - 《R 语言数据分析实战》 <https://xiangyunhuang.github.io/data-analysis-in-action/> 本书围绕数据分析实战工作流分四大部分：数据收集和整理，数据探索和分析，数据建模和解释，数据交流和应用。本书的 R 语言的工具有介绍与数据分析的实战案例并重，涵盖网络数据、空间数据、时序数据和文本数据。
 - 《现代统计图形》第二版 <https://xiangyunhuang.github.io/msg2nd/> 本书是对第一版的重写，更加注重图形功能分类、使用场景和绘图工具的介绍，这是对第一版最大的补充。
 - 《现代应用统计》 <https://masr.netlify.app/> 本书分三大部分，分别是机器学习、贝叶斯建模和空间分析。三个部分分别依据任务类型、模型种类和数据类型展开。应用机器学习首先需要确定任务类型，根据不同的任务选用不同的算法。机器学习根据任务类型分聚类、分类、回归和排序四个章节。
 - 在个人博客上累计创作了 110 篇，其中技术博客 40 篇、个人日志 55 篇，财经商业 15 篇。



联系方式

✉ xiangyunfaith@outlook.com
🌐 github.com/XiangyunHuang

工具技能

专题数据分析、数据异动归因、BI 报表开发、数据可视化等方面经验丰富。

熟练掌握 R 语言、SQL、Git，熟悉 Spark、Hive，了解 Presto、MySQL、Doris 和 Bash。

熟练掌握 R Markdown / R Shiny / ggplot2 / plotly 等工具。

2020.10
|
2022.04

数据研发工程师 (商分方向)

美团, 美团平台, 数据智能组

北京, 中国

- 围绕搜索业务和搜索产品, 参与搭建和维护搜索数据指标体系, 在部门核心离线指标的异动分析方面, 协同引导团队、DQU 团队、产品团队 (数据产品、业务产品、埋点产品)、运营团队、大数据团队, 做了很多降本提效、数据驱动业务的工作。
- 搜索渗透率: 分用户属性、业务类型、多级渠道建立核心看板工具, 持续观测部门核心指标, 在业务异动分析, 确定团队业绩指标等方面起到重要作用。
- 城市大盘: 搜索是连接用户需求和门店供给的重要方式, 用户需求受运营活动、热点突发、节假日、季节、气温、政策等诸多因素影响, 分析团队、产品团队、运营团队在归因分析方面压力巨大, 提取通用的Query、时段、业务、城市等维度, 从空间分布、趋势变化、占比变化和对比差异等多角度呈现部门核心业务指标, 在搜索运营、热点突发、指标异动等方面可以快速归因, 已作为搜索运营团队的日常工具。
- 业务 Query 分析: 任选时间段和业务, 圈选出Top Query、涨跌词、新增词和减少词, 发现各个业务下用户需求的变化, 为搜索运营、业务团队提供洞察工具, 除搜索技术部外, 工具助力十余个业务部门。
- 搜索换金刚: 从 Query、意图和业务等维度分析用户从美团大搜跳出进入金刚下单消费的行为, 并工具化, 在搜索 Badcase 发现、供给接入, 以及后续的分析和改进等方面持续助力, 驱动搜索运营、业务运营、算法策略优化。理想态下可为搜索业务每日带来数百万的业绩。

2019.10
|
2020.09

数据分析师

51Talk, 大数据智能中心, 分析组

北京, 中国

- A/B 测试: 学术线 (体验课、公开课和常规课) 教材、产品和课研相关的 A/B 测试项目, 参与建设 A/B 测试平台, 伴随 A/B 测试的专项分析报告等
- 指标体系: 公司级组织绩效看板开发, 包括梳理和组织指标体系 (涉及指标定义、内容和计算), 完成底层数据到业务指标的计算, 规划和交付页面设计内容。
- 技能培训: 上海、武汉分析师培训, 部分团队新人和学术产品, 覆盖 R 语言数据操作、Git 代码管理、Azkaban 任务调度、RStudio 集成环境使用等。
- 报表开发: 菲教运营报表开发 (包含体验课和常规课)、青少付费学员画像、课耗日活跃度、体验课教材转化率、公开课 KPI 等报表开发。

2019.01
|
2019.03

软件工程师

文因互联

北京, 中国

- 利用 Python 和正则表达式从上证披露的大量 PDF 文档中提取信息, 参与互联网金融问答、搜索引擎的开发, 以自然语言处理技术构建规则系统, 基于大规模正则表达式的开发。

📁 实习经历

2017.09
|
2017.11

● 数据分析师

新浪, 新浪新闻, 运维技术中心

📍 北京, 中国

- 负责新浪新闻客户端日志数据的分析工作, 支持业务部门策略研究、服务器资源调度、查找资源访问的瓶颈及原因、服务器运行状态的实时监控。
- 在部门首次独立将 R 语言引入大规模日志分析, 借助 Docker、ClickHouse 和 R 制作自动化日报。相比于之前基于 SQL/Excel 的工具链大大提高工作效率, 使得数据分析 workflow 更加稳健, 自动化分析 APP 性能。
- 在新浪 BIP 平台部介绍如何将 R 语言引入日常报表的开发, 完成从数据源到报表的一键生成和自动化, 在 2018 年中国 R 语言会议上给报告介绍背后的技术。
- 基于数据仓库 ClickHouse, 容器 Docker 和编程语言 R (包括 knitr/DBI/ggplot2/gganimate 等包) 制作中文 R Markdown 模板, 同时生成 PDF 和 HTML 格式的数据分析报告, 并添加定时任务。
- 在 R Markdown 模板的基础上自动化生成常规报表和可视化图, 缓解每次启动 superset 做可视化报表而查询数据库的压力, 同时节省时间。
- 在 R 包的基础上继续开发, 添加 API 接口的技术文档, 含参数的 SQL 查询结合 knitr 缓存的介绍, 添加更多实际数据的使用案例。

📝 开源项目

2016 -

● 🏠 个人博客 <https://xiangyun.rbind.io/>

我的主页开辟了博客、日志、商业三个栏目。博客栏目介绍工具、算法、模型等技术性内容, 比如数据可视化、可重复性文档、空间统计、数据分析等。日志栏目写一些个人所见、所闻、所思。商业栏目围绕工作、生活中的现象思考背后的商业模式, 比如互联网公司的概况、短视频和直播背后的商业模式等。

📍 北京, 中国

2019 -

● 📖 学习笔记 <https://notesdown.netlify.app/>

数据读写、数据操作、数据可视化、数值优化等方面的技术笔记

📍 北京, 中国

📖 出版专著

2021

● 《现代统计图形》 <https://book.douban.com/subject/35547494/>

人民邮电出版社

📍 北京, 中国

赵鹏, 谢益辉, 黄湘云.



荣誉奖项

2011.09



2018.12

中国矿业大学（北京）

理学院

📍 北京, 中国

- 2018年 RStudio 第一届 bookdown 大赛 (亚军奖)
- 2016—2017学年优秀研究生荣誉称号 (一等奖)
- 2015—2016学年研究生优秀学生 (一等奖)
- 2013年高教社杯全国大学生数学建模竞赛 (二等奖)
- 2013年校物理实验竞赛 (一等奖)
- 第29届全国部分地区物理竞赛 (三等奖)
- 2012—2013学年本科生优秀学生奖学金 (二等奖)
- 2013年校科技文化节科普创作大赛 (优秀奖)
- 全国大学英语四级和六级