# 数据研究部管理制度

# 第一部分、总则

为保证数据研究部日常工作正常有序地进行，让各个工作环节更加紧凑、可控，需要尽可能地实现数据工作的规范化，工作流程的标准化，以提高工作质量和工作效率，便于更好地完成部门任务。

## 第二部分、部门文化

## 1. 核心价值观

责任心 严谨 认真

在数据研究部的核心价值观中，责任心是位列首位的，因为数据工作是非常严谨的，作为数据研究人员必须时刻将责任记在心上，无论工作难易都应该以认真负责的态度去执行。

## 2. 部门定位

数据支持 研究探索

数据研究部在公司中的主要作用是为公司的税务业务及网站运营提供数据支持，利用数据研究为公司及部门探索新的潜在业务。

## 3. 部门目标

精准无误 开拓创新

对于日常的数据工作，需要部门成员严谨细心地完成，力求做到零失误。数据研究工作要摆脱桎梏，推陈出新。

## 4. 工作态度

耐心 细心 责任心

数据工作对公司非常重要，容不得一丝马虎，某次不小心的失误往往会在将来的工作中造成重大的损失，作为数据研究部的员工，日常工作要力求完美。

# 第三部分、部门任务

根据公司现状及公司的未来发展规划，为了提高存量房网格化价格数据业务的效益、配合人与网的运营与扩展，数据研究部主要有如下几项任务：

## 1. 数据生产

数据生产任务源于安徽省各地市地税系统存量房数据更新项目，具体的数据更新量视项目合同而定，数据生产由市场人员传达数据生产任务，由部门负责人安排数据专员负责具体工作。数据生产的常规工作流程如下：

### （1）制作数据采集模板

数据采集模板是用于评估公司（或其他数据采集机构）采集可比实例数据之用，数据更新项目合同签订后，数据专员需对数据采集模板进行维护，确认无误后，由网络部将采集模板挂载在人与网，由地税部门专人登陆下载。

### （2）人工审查

地税部门将评估公司采集的数据上传之后，由部门负责人下载并分配给数据专员进行人工审核。

若采集数据不符合数据采集模板要求，则编写问题反馈文档，并对问题数据进行标记，交由部门负责人上传至人与网，并通知地税部门专人查收，有地税部门安排评估公司对采集数据进行修正。

### （3）数据转码及报告编写

确认采集数据无误后，由数据专员根据采集数据编写数据更新总结报告、对采集数据进行转码，并将报告与数据文件交由部门负责人上传至人与网。

上传后通知省局值守人员查收转码后的数据、通知地税部门查收数据更新总结报告。

### （4）对数据进行灌装入库

省局值守人员下载数据后，灌装入测试数据库并对数据进行测试，如果测试有误则返回数据研究部进行修改，直至测试无误为止。

测试无误后，由值守人员将数据灌装入正式数据库，并对旧数据进行备份。

## 2. 数据采集

在价格模型研究前需要采集大量数据，采集到的数据除用于价格模型研究外，还可以提供给人与网做存量房数据展示或为其他模型相关的评估产品提供数据支持。

以存量房模型为例，数据采集前期工作由数据专员负责，包括学区、土地级别、大型商超、医院等数据的信息采集、地址定位和人工分级。

数据采集后期工作由数据研究部研究员负责，主要针对模型进行数据补充，帮助扩大不同存量房之间的特征差异，从而强化模型评估效果。

### （1）学区数据和土地级别数据采集及绘制

根据各地教育局公布的学区范围和国土局公布的土地级别数据，在公司自制的网格化平台进行绘制。目前安徽省市级学区和土地级别都已经绘制完成，后续的工作主要是县级学区和土地级别的绘制。

### （2）配套设施数据采集及入库

通过电子地图、医疗健康网站、公交查询网站等渠道，通过爬虫或公开API进行数据采集（医院、商超、银行、学校、地标、负面因素等），在数据采集完成之后，由数据专员对这些数据进行分级及数据格式整理，将整理好的数据交由网格化数据管理员存入网格化数据库。

### （3）因素补充

在价格模型初步建立后，如果模型效果不理想，则需根据该地市的实际情况进行分析，查找模型误差的来源，从而有针对性地对特定因素进行补充，补充完成之后重新训练模型，查看效果，知道模型效果达标位置，此项工作由数据研究员完成。

### （4）新模型研究及标准文档的形成

为了人与网评估产品的业务拓展，数据研究部需要投入较多的时间研究新模型，如二手车价格评估模型。

每个模型研究完成之后，都需要编写标准文件，协助标准研发部完善标准内容，依靠标准来提高评估产品的权威性和科学性。

## 3. 价格模型研究

价格模型研究由数据研究部研究员负责。以下工作内容均以目前进行的存量房价格评估模型为例，日后若研究其他模型，工作内容会稍有变动。

### （1）因素与片区、小区的匹配

模型中评价的主体是片区和小区，所以需要将现有的区位与周边配套因素转化为片区和小区的特征，才能对评估的价格产生影响。

### （2）特征整理

匹配好的特征不一定符合模型的要求，可能需要进行进一步的整理，整理工作包括：数值化、连续数据离散化、独热编码等

### （3）模型训练

模型训练最重要的是选择合适的算法，根据不同的需求选择合适的算法，例如需要形成文档的部分尽量采用直观的模型如多元线性回归、决策树等，而追求准确率而不要求特征解释的部分可以尝试SVM和神经网络等黑箱模型。

### （4）模型验证

在训练之初需要将数据均匀划分出训练集和测试集，模型训练完成之后，在测试集上进行第一次验证，如果验证效果不好则返回修改特征或者算法，直到模型预测结果合格为止。

第一次验证完成之后，采用新数据继续进行验证，同样，效果不好则返回修改，直到模型预测结果合格为止。

## 4. 数据分析

数据研究部将在未来为人与网提供数据分析技术支持，为人与网的发展、运营方向提供切实可行的建议。为了了解人与网的运行现状和可行的优化方案，需要进行如下五项工作：

### （1）流量分析

流量是最基本也是最重要的网站数据分析维度，它反映了一个网站被访问的整体概况。流量分析包含两方面内容：流量趋势分析和当前实时流量。通过网站的流量趋势分析可以对当前网站的运营策略的效果进行评估。

### （2）来源分析

来源分析是对网站流量的来源渠道进行统计分析，它可以反映访客是通过何种渠道访问网站的。流量来源一般可以分为三大类，直接流量、推介流量和搜索引擎流量。根据流量来源来决定网站的优化策略，

### （3）页面分析

页面分析是对访客访问网站各个页面的频次、时长等信息进行统计分析，从而找出高频使用的页面或功能，掌握访客兴趣和热点页面，另外，还可以通过分析页面的访问路径，获取访客的操作习惯。页面分析是优化页面内容和页面间逻辑流程的重要依据。

### （4）用户画像建立

建立用户画像的目的是通过分析用户的行为，为每个用户打上标签并计算该标签的权重，需要对用户数据进行采集，用户数据包括动态数据（浏览习惯、偏好）和静态数据（人口属性、商业属性）。

### （5）用户喜好推荐系统建立

根据用户画像搭建智能推荐系统，为新老用户推荐适合他的或者他可能感兴趣的评估产品，帮助吸引新客户留住老客户，并提升用户在人与网的消费倾向。

## 第四部分、岗位职责

数据研究部员工分为数据专员与数据研究员：

1. 数据专员

数据分析师主要负责数据生产和网格化采集工作。

|  |  |
| --- | --- |
| 部门任务 | 细分职责 |
| 数据生产 | 1. 可比实例采集模板的制作；  2. 地税局可比实例采集数据审核；  3. 可比实例数据转码；  4. 可比实例数据入库；  5. 数据问题处理；  6. 存量房数据日常维护。 |
| 数据采集 | 1. 网格化平台各地区学区和土地级别数据采集和绘制；  2. 各地区小区配套设施（医院、商超、银行、学校、公交、地标、负面因素等）采集、定位及分级；  3. 小区地址、配套设施、开发商、物业信息核对；  4. 网格化数据日常维护。 |
| 价格模型研究 | 配合数据研究院进行网格化数据的人工审查。 |
| 数据分析 | 1. 人与网日常流量分析；  2. 人与网定期流量来源分析；  3. 人与网定期页面分析。 |

2. 数据研究员

数据研究院主要负责存量房价格模型研究和网格化平台数据补充工作。

|  |  |
| --- | --- |
| 部门任务 | 细分职责 |
| 数据生产 | 为数据专员进行数据分析及公式校正等技术支持。 |
| 数据采集 | 1. 就需要采集的数据类型、数量、分级标准等要求与数据专员做详细交流；  2. 在价格模型准确度不够时，通过技术手段进行特征补充。 |
| 价格模型研究 | 1. 根据地理坐标将小区周边的配套设施和区域信息与小区进行匹配；  2. 特征工程：即将小区特征转化成模型可用的形式；  3. 小区区位特征与价格的关系模型搭建；  4. 研究新模型；  5. 模型技术文档的编写 |
| 数据分析 | 1. 用户画像建立；  2. 智能推荐系统算法研究。 |
| 其他 | 人与网大数据简报维护 |

## 第五部分、作业指导书

1. 数据生产过程作业指导见《数据研究部——数据生产作业指导书》；

2. 数据采集过程作业指导见《数据研究部——网格化工作平台作业指导书》；

3. 价格模型研究过程作业指导见《数据研究部——价格模型研究作业指导书》。

## 第六部分、部门制度

### 1. 数据保密

工作中接触到的存量房数据信息必须严格保密，严禁外传泄露，若发生数据泄露公司将有权对泄露这追究法律责任；

### 2. 考勤

员工考勤和请假事宜严格遵守《员工手册》进行；

### 3. 工作安排

（一）数据生产工作根据项目实际需要，由部门负责人进行工作分配，员工应服从工作任务分配，并对分配到的工作总结和报告应及时完成。负责采集的各种信息也应及时沟通汇总；

（二）员工应定期对自己的工作进行记录，记录内容包括工作的过程、结果、工时及使用的方法，便于月末的工作汇报；

（三）每个月末，员工应将本月完成的主要工作及重点工作记录整理好，发送给部门负责人，负责人将本月部门重点工作记录进行汇总整理，向上级汇报；

（四）部门在每月末召开工作计划会议，对本月的工作计划完成情况进行总结汇报，指定下一个月的工作计划，包括工作内容、计划完成时间、各人职责等。

### 4. 团队协作

（一）如果在个人工作上出现难以解决的问题，部门员工需要互相帮助、共同分析，找出问题根源所在，并在解决问题之后对该问题的解决过程和解决方案进行记录，避免以后发生类似的情况。

（二）部门员工必须积极主动地为部门建设出谋划策，有问题要积极反映，将对数据研究部的想法与意见全面客观的反映出来，真正的参与到部门的工作中去；

## 第七部分、激励措施

为了提高各工作岗位的工作积极性和竞争力，客观准确地评价员工工作职责和任务完成情况，特制定本部门激励管理办法。

### 1. 月度绩效考核

（一）考核计划

数据研究部所有员工参与考核，试用期内的员工一同进行考核但绩效工资固定不变；每月最后一周制定下月数据研究部KPI计划，每月第一周计算数据研究部以及所有员工绩效级别，每个自然月为一个考核周期。

（二）工资组成

员工工资 = 基本工资（工资总额70%） + 绩效工资（工资总额30%）

（三）绩效计算

管理人员绩效得分 = 数据研究部绩效得分

普通员工绩效得分 = 员工KPI得分（70%） + 员工日常工作得分（20%） + 上级评分（10%）

（若当月部门有重大里程碑事件，则员工日常工作得分将替换为该里程碑得分）

普通员工日常工作得分：根据员工日常工作的完成率，上级评价得分；

上级评分：根据员工日常工作总结和计划，工作态度，工作难度，工作质量，执行力，责任心，协作沟通，创新力等进行评分。

（四）绩效工资计算

|  |  |  |
| --- | --- | --- |
| 绩效得分 | 绩效级别 | 绩效工资 |
| 100 - ∞ | A | 150% |
| 95 - 100 | B | 120% |
| 75 - 95 | C | 100% |
| 60 - 75 | D | 80% |
| 0 - 60 | E | 50% |

（60对应D；75对应C；95对应B；100对应A）

### 2. 年终奖金分配

（一）部门可分配年终奖总额由公司根据《公司18年度存量房网格化价格数据业务的激励协议》确定；

（二）年终根据部门职工的月度累积KPI考核情况进行年终奖金分配，奖金按如下方法计算:

个人奖金 = 部门可分配奖金×岗位系数×个人绩效得分/部门员工绩效总分