# 数据研究部管理制度

# 第一部分、总则

为保证日常工作正常有序地进行，让数据研发部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、部门任务

根据公司现状及发展情况，为了开拓公司业务，稳固人与网的运营，数据研究部主要有如下几项任务：

## 1. 数据生产

数据生产任务源于地税系统存量房数据更新项目，工作量视项目合同而定，数据生产由数据业务人员传达数据生产任务，由数据研究部负责人安排数据专员负责具体工作。

### （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，通知地税部门专人下载。

### （2）人工审查

此项工作由数据专员完成，若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。

### （3）数据转码及报告编写

确认数据采集数据无误后，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，并通知省局值守人员查收、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，可以为人与网的评估业务提供助力。以存量房模型为例，数据采集前期工作由数据研究部数据专员负责，包括学区、土地级别、大型商超、医院等的采集和分级。

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。

### （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。

### （2）配套设施数据采集及入库

通过电子地图、医疗网站、公交网站等渠道进行数据采集（医院、商超、银行、学校、地标、负面因素等），并对这些数据进行分级，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，对特定城市可能有特定因素需要补充，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

为了公司的长远发展，需要投入较多的时间研究新模型，如二手车价格评估模型。对于每个模型都需要形成标准文件，以提高其权威性，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

评价的主体是片区和小区，所以需要将现有的数据转化为片区和小区的特征，才能应用于模型当中。

### （2）特征整理

匹配好的特征往往不符合模型的要求，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，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

在人与网用户量增大到一定量级后，数据研究部将为人与网提供数据分析技术支持。

### （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。

### （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。

### （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。

## 第三部分、部门文化

<https://wenku.baidu.com/view/0d8d7418561252d380eb6ed2.html?from=search>

## 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； |
| 数据采集 | 1. 网格化平台小区定位人工审核；  2. 网格化平台各地区学区和土地级别数据采集和绘制；  3. 各地区小区配套设施（医院、商超、银行、学校、地标、负面因素等）采集、定位及分级。 |
| 价格模型研究 | 配合数据研究院进行数据检查 |
| 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 |
| 其他 |  |

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 |
| 岗位任务 | 数据模型研究，技术文档编写，模型算法研究 |
| 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

## 第六部分、部门制度

工作制度

1. 工作中接触到的隐私信息必须严格保密，严禁外传泄露；

2. 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；

3. 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；

4. 部门员工不得无故旷工，如遇特殊情况，应提前向部门经理请假。

办公室制度

1. 部门职工每周轮岗之日，负责办公室卫生；

2. 每天下班最后一个离开办公室的员工需管好门窗；

## 第七部分、激励措施