景气指数模型建设需求

一、建设目标

建立投资景气指数的计算模型，并实现自动化的数据更新、模型计算和大屏展示。

二、功能性需求

（一）数据更新与抽取

1、定期从指定共享库中读取数据，并按照指定聚合方式形成时间序列数据；

2、涉及到两部分数据，一是投资审核备数据，共两张表，数据表和字典表，数据表内容如下：

|  |  |  |  |
| --- | --- | --- | --- |
| **字段说明** | **字段名称** | **字段类型** | **备注** |
| 项目编码 | project\_code | varchar |  |
| 项目名称 | project\_name | varchar |  |
| 大区域 | area | varchar | 字典表中group\_no=1008 |
| 区县 | county | varchar | 字典表中group\_no=1008 |
| 父级大行业 | super\_parent\_Industry | varchar | 字典表中group\_no=20210305 |
| 大行业 | parent\_Industry | varchar | 字典表中group\_no=20210305 |
| 小行业 | child\_Industry | varchar | 字典表中group\_no=2 |
| 投资类型 | investment\_type | varchar | 0-内资，1-外商投资，2-境外投资 |
| 项目建设类型 | build\_type | varchar | 字典表中group\_no=1009 |
| 资金来源 | invest\_source | varchar | 1-政府投资，2-企业投资 |
| 实际开工时间 | actual\_Start\_Time | date |  |
| 实际完工时间 | actual\_End\_Time | date |  |
| 预计开工时间 | expect\_start\_time | date |  |
| 预计完工时间 | expect\_end\_time | date |  |
| 投资总金额 | total\_investment | decimal | （单位：万元） |
| 自有投资 | qytzzj | decimal | （单位：万元） |
| 银行贷款 | qydkzj | decimal | （单位：万元） |
| 审批-中央财政性投资（小计） | zyczxtz | decimal | （单位：万元） |
| 审批-中央预算内投资 | zytzzj | decimal | （单位：万元） |
| 审批-中央预算内专项资金 | zyzxzj | decimal | （单位：万元） |
| 审批-中央专项建设基金 | zyjszj | decimal | （单位：万元） |
| 审批-其他中央财政性投资 | zyqtzj | decimal | （单位：万元） |
| 审批-地方财政性投资（小计） | dfzjhj | decimal | （单位：万元） |
| 审批-市预算内统筹资金 | sjtczj | decimal | （单位：万元） |
| 审批-市财政专项资金 | sjzxzj | decimal | （单位：万元） |
| 审批-市级部门自筹投资 | sjzczj | decimal | （单位：万元） |
| 审批-其他地方财政性投资 | sjqtzj | decimal | （单位：万元） |
| 审批-利用外资（小计） | wszjhj | decimal | （单位：万元） |
| 审批-外商直接投资 | wstzzj | decimal | （单位：万元） |
| 审批-主权外债 | wszqzj | decimal | （单位：万元） |
| 审批-外国政府赠款 | wsjzzj | decimal | （单位：万元） |
| 审批-企业投资（小计） | qyzjhj | decimal | （单位：万元） |
| 企业其他渠道融资 | qyrzzj | decimal | （单位：万元） |
| 区县自筹资金 | qxzczj | decimal | （单位：万元） |
| 项目单位自筹资金 | xmdwzczz | decimal | （单位：万元） |
| 借用外国政府贷款 | jywgzfdk | decimal | （单位：万元） |
| 其他投资（小计） | qtzjhj | decimal | （单位：万元） |
| 审批项目类型 | approve\_type | varchar | 1:审批;2:核准;3:备案;4:项目建议书(立项);5:可行性研究报告;6:初步设计及概算;7:其他 |
| 审核备项目创建年月 | shb\_create\_month | varchar |  |
| 审核备办结大部门 | shb\_bj\_parent\_department | varchar | 字典表中group\_no=20210304 |
| 审核备办结小部门 | shb\_bj\_child\_department | varchar | 字典表中group\_no=20210304 |
| 审核备办结时间 | shb\_bj\_time | date |  |
| 审核备办结年 | shb\_bj\_year | int |  |
| 审核备办结月 | shb\_bj\_month | int |  |
| 审核备办结年月 | shb\_bj\_year\_month | int |  |
| 经济类型 | economic\_type | int | 1:国家及国有控股经济;2:集体经济;3:私营经济;4:股份制;5:其他 |
| 创建时间 | create\_time | timestamp |  |
| 更新时间 | update\_time | timestamp |  |
| 版本号 | version | int |  |

**建议将表聚合的方式，写成配置文件的形式**。

二是统计数据，包含主要经济运行指标数据，从共享库中读取。

3、要有页面可以看到数据抽取的状态，成功与否。有“立即更新”按钮。

（二）模型建设与计算

1、实现景气指数的计算模型，可使用Python、R、MATLAB或Java实现，也可尝试直接调用spss、stata等专业软件

2、景气指数模型的步骤主要分为：

（1）初选指标（从宏观经济指标、投资情况指标、政务大数据指标（审核备指标、信用指标等）、企业指标（如利润率等）等几个维度进行考虑）

（2）数据预处理（包括缺失值填充（如某些指标缺少1月数据）、数据频度转换、数据类型转换、定基价格计算、季节调整（剔除季节因素[[1]](#footnote-1)和不规则因素）等）

（3）指标遴选（先行、一致、滞后，可采用时差相关系数和K-L信息量法确定滞后期数）

（4）指标精选（结合实际操作及数据质量情况最终确定）

（5）合成指数计算[[2]](#footnote-2)（形成先行合成指数、一致合成指数、滞后合成指数）

①计算指标对称变化率

②将指标对称变化率进行标准化处理

③计算各指标组的标准化平均变化率

④计算初始合成指数

⑤进行趋势调整

⑥设定合成过程权重

⑦计算三个指标 组合成指数

（6）投资景气预警

①确定预警指标

②确定预警界线值

③计算综合警情指数并确定其界限值（红黄蓝信号灯）

④分析投资处于运行状态及其原因

3、实现两种计算方法的景气指数模型，分别是皮尔逊相关系数[[3]](#footnote-3)和K-L信息散度[[4]](#footnote-4)，根据计算方法计算基准指标与被选择指标（含时差偏移）之间的关系，从而筛选出先行指数和一致指数。

4、功能上，既要保持模型计算的自动化，也要可以对必要步骤、必要参数进行审核和人工介入。比如后台有模型配置的页面，可以选择或修改基准指标、被选指标，可以选择不同的计算方法。

（三）大屏展示

单页面展示景气分析的结果，主要包括两部分：一是用曲线图表示景气指数和投资完成情况的走势；二是用信号灯表示各景气指标的预警情况。

三、非功能性需求

1、采用模块化、低耦合的架构，模型可通过接口或者数据库与系统交互，模型可替换。

2、定期自动化计算，计算实时性要求不高；

3、发改委目前信息化的环境主要是deepin server和达梦数据库，开发尽量遵循以上规范。

四、时间安排

4月20日，项目启动；

4月21日，环境搭建，数据准备；

4月22日-4月23日，完成数据更新与抽取开发；

4月22日-4月29日，完成模型建设与计算开发；

4月23日-4月29日，完成大屏展示设计与开发；

4月30日，测试与初步成果演示；

5月6日-5月7日，完善成果。

1. 参考：Python中有现成库“seasonal\_decompose”可对数据进行季节性分析，已测试调通 [↑](#footnote-ref-1)
2. 有成熟的合成方法，可沟通 [↑](#footnote-ref-2)
3. 皮尔逊相关系数比较简单，可直接参考百科 [↑](#footnote-ref-3)
4. 可参考维基百科和附件，Python中“scipy”有现成方法，且自动进行归一化处理。 [↑](#footnote-ref-4)