清洗规范

目标字段和垃圾字段用@\_@分隔

# 爬虫招标信息原始表

ti\_tender\_detail\_day

## tender\_code

唯一UUID

## prjct\_name

非空 非仅由数字，符号组成的字符串

其中包含项目名称字段，若包含其他数据需要有特定的区分

例：关于杭州市交通警察支队文教中队指挥中心建设项目的采购公告

## prjct\_code

非空

其中包含项目编号字段，若包含其他数据需要有特定的区分

例：XHZFCG-2019-G-01

## prjct\_desc

项目描述字段

例：根据《中华人民共和国政府采购法》、《政府采购货物和服务招标投标管理办法》等规定，经杭州市余杭区财政局确认书余财采确临[2018]4895号批准，现就杭州市公安局余杭区分局交通警察大队的迎宾大道北延隧道交通工程保安服务项目进行公开招标。欢迎符合要求并有能力完成本项目的投标人前来投标。

## prvnce\_name

非空

所属标准地址的省份

其中包含该字段，若包含其他数据需要有特定的区分

例：浙江

## latn\_name

非空

所属标准地址的地市

其中包含该字段，若包含其他数据需要有特定的区分

例：杭州

## county\_name

非空

所属标准地址的区县

其中包含该字段，若包含其他数据需要有特定的区分

例：下城区

## release\_time

非空

发布时间 格式：YYYY-MM-DD

其中包含该字段，若包含其他数据需要有特定的区分

例：2019-05-10

## begin\_time

非空

开始时间 格式：YYYY-MM-DD

其中包含该字段，若包含其他数据需要有特定的区分

例：2019-05-10

## end\_time

非空

截止时间 格式：YYYY-MM-DD

其中包含该字段，若包含其他数据需要有特定的区分

例：2019-05-10

## tender\_unit

非空

非仅由数字，符号组成的字符串

招标单位字段

其中包含该字段，若包含其他数据需要有特定的区分

例：杭州市公安局余杭区分局交通警察大队

## contactor

招标人字段

仅由中文组成

其中包含该字段，若包含其他数据需要有特定的区分

例：郭振祥

## contact\_phone

招标人联系方式字段

仅由数字组成

其中包含该字段，若包含其他数据需要有特定的区分

例：0571-86238495

## purchase\_money

非空

金额字段 取预算金额

带上单位 大写金额或者数字金额取一种

其中包含该字段，若包含其他数据需要有特定的区分

非分包情况下写入该字段。purchase\_money\_all字段为空

例：1950000

## agent\_unit

采购代理单位

其中包含该字段，若包含其他数据需要有特定的区分

例：山东东岳项目管理有限公司

## agent\_contactor

代理联系人

其中包含该字段，若包含其他数据需要有特定的区分

例：张飞

## agent\_phone

代理联系方式

仅由有效数字组成

例：0411-1111222

其中包含该字段，若包含其他数据需要有特定的区分

## bidder\_req

## tender\_note

## open\_note

以上字段内容 去掉标点符号。

## inter\_name

网站名称

## website

详情页链接

## tender\_detail

## etl\_time

处理时间

## purchase\_money\_all

分包金额，用@\_@分隔

存在分包情况下，写入该字段。purchase\_money为空

## money\_num

存在分包情况下，记录分包个数

# 爬虫中标信息原始表

ti\_winbidder\_detail\_day

## winbidder\_code

唯一UUID

## prjct\_name

非空

其中包含项目名称字段，若包含其他数据需要有特定的区分

例：关于杭州市交通警察支队文教中队指挥中心建设项目的采购公告

## prjct\_code

非空

其中包含项目编号字段，若包含其他数据需要有特定的区分

例：XHZFCG-2019-G-01

## prvnce\_name

非空

所属标准地址的省份

其中包含该字段，若包含其他数据需要有特定的区分

例：浙江

## latn\_name

非空

所属标准地址的地市

其中包含该字段，若包含其他数据需要有特定的区分

例：杭州

## county\_name

非空

所属标准地址的区县

其中包含该字段，若包含其他数据需要有特定的区分

例：下城区

## release\_time

非空

发布时间 格式：YYYY-MM-DD

其中包含该字段，若包含其他数据需要有特定的区分

例：2019-05-10

## begin\_time

非空

开始时间 格式：YYYY-MM-DD

其中包含该字段，若包含其他数据需要有特定的区分

例：2019-05-10

## tender\_unit

非空

非仅由数字，符号组成的字符串

招标单位字段

其中包含该字段，若包含其他数据需要有特定的区分

例：杭州市公安局余杭区分局交通警察大队

## contactor

招标人字段

仅由中文组成

其中包含该字段，若包含其他数据需要有特定的区分

例：郭振祥

## contact\_phone

招标人联系方式字段

仅由数字组成

其中包含该字段，若包含其他数据需要有特定的区分

例：0571-86238495

## agent\_unit

采购代理单位

其中包含该字段，若包含其他数据需要有特定的区分

例：山东东岳项目管理有限公司

## agent\_contactor

代理联系人

其中包含该字段，若包含其他数据需要有特定的区分

例：张飞

## agent\_phone

代理联系方式

仅由有效数字组成

例：0411-1111222

其中包含该字段，若包含其他数据需要有特定的区分

## winbidder\_unit

中标单位

要求参考tender\_unit

若中标存在联合中标的情况下，扩展字段

## winbidder\_money

中标金额

非联合中标情况下，中标金额写入该字段。·

例： 111122244

## bid\_time

定标/开标 日期

格式 YYYY-MM-DD

## bid\_month

定标/开标 月份

YYYY-MM

## inter\_name

网站名称

## website

详情页链接

## winbidder\_detail

## etl\_time

处理时间

## winbidder\_unit\_all

联合中标单位，用@\_@分隔

存在联合中标的情况，写入该字段

## winbidder\_money\_all

联合中标金额，用@\_@分隔

存在联合中标的情况，写入该字段

## money\_num

存在联合中标的情况下，记录联合中标个数