

README

安装：使用python uv 安装所有依赖

```
uv sync
```

配置：

- 1. setting.toml 文件：REWORKSPACE 这个配置到本地目录
- 2. 码运行主要在： `test_standard_process.py` 这个文件

代码运行步骤:

- 1. `test_download_all_bonds` 这个函数用来下载pdf文件 修改年份范围和省份范围，可以指定某个省某个年份范围下载PDF,下载完成之后的文件目录结构：{REWORKSPACE}/<省份>/<年份>

```
def test_download_all_bonds():
    for year in range(2024, 2014, -1):
        # provinces = get_province_list(["广东省", "黑龙江省", "山东省", "福建省", "辽宁省", "云南省", "吉林省", "四川
        # DONE_PROVINCE = [""]
        ## 西藏自治区, 海南省, 青海省, 宁夏回族自治区, 江苏省, 陕西省
        # DONE_provinces = ["吉林省", "云南省", "四川省", "内蒙古自治区"]
        # provinces = ["宁夏回族自治区", "安徽省", "山东省", "山西省"]
        # provinces = ["西藏自治区", "江苏省", "浙江省", "河北省"]
        # -----
        # provinces = ["河南省", "浙江省", "海南省", "湖北省"]
        # provinces = ["湖南省", "甘肃省", "福建省", "西藏自治区"]
        # provinces = ["贵州省", "辽宁省", "陕西省", "青海省"]
        provinces = ["陕西省", "安徽省"]
        ## 河南省/山东省/福建省
        for province in provinces:
            set_parameters(str(year), province)
            pd_ext.filter_table_file(Path(REWORKSPACE + "/overall.csv"), col_filters,
                                     REWORKSPACE + "/" + filtered_file_name)
            os.makedirs(download_dir, exist_ok=True)
            download_bonds(source_file_path, download_dir)
```

2`test_unzip_all_files` 这个用来unzip下载的所有zip包里面的文件到指定目录all_files目录,修改省份和年份就可以

```
def test_import_project_detail_data():
    """
    1. manual check and collect first raw data
    2. cleanup data for csv project
    :return:
    """
    set_parameters( year: "2015", province: "宁夏回族自治区")
    cleanup_source_file_path = cleaned_file
    cleaned_df = pd_ext.read_to_df(cleanup_source_file_path)

    models = pd_ext.read_to_models_by_field_mapping(cleaned_df, model_class=BondProjectDetail
                                                    , field_mappings=FIELD_MAPPINGS)

    for model in models:
        try:
            value = csv_ext.replace_dots_except_last(model.amount)
            # if float(value)>1000:
            model.amount = float(value) / 10000
        except Exception as e:
            model.amount = model.amount
        db.save(model)
```

3test_import_project_detail_data 用来将csv数据清洗，结构化，导入数据库，修改年份和省份，从 <年份>-<省份名>.csv 文件中读取数据，清洗结果的文件是 cleanup-<年份>-<省份名>.csv

```
def test_import_project_detail_data():
    """
    1. manual check and collect first raw data
    2. cleanup data for csv project
    :return:
    """
    set_parameters( year: "2015", province: "宁夏回族自治区")
    cleanup_source_file_path = cleaned_file
    cleaned_df = pd_ext.read_to_df(cleanup_source_file_path)

    models = pd_ext.read_to_models_by_field_mapping(cleaned_df, model_class=BondProjectDetail
                                                    , field_mappings=FIELD_MAPPINGS)

    for model in models:
        try:
            value = csv_ext.replace_dots_except_last(model.amount)
            # if float(value)>1000:
            model.amount = float(value) / 10000
        except Exception as e:
            model.amount = model.amount
        db.save(model)
```

≡ 2023-四川省.csv
📄 2023-四川省.xlsx
≡ cleanup-2020-四川省.csv

4. ``test_cleanup_import`` 直接从清洗过的文件中导入数据库,使用基本同第三步骤，主要发现有些原始文件保存数据库出错，然后修改一下清洗文件之后，直接导入

数据从哪里来

pdf文件下载完成之后,

1. 通过WPS 的 **PDF转换成EXCEL功能** 批量转换PDF成EXCEL文件

- | 名称 | 项目名称 | 地区 | 项目总投资 | 项目拟发额 |
|-----------------------------------|------------------------|--------|-------|-------|
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 白城粮食产业园示范区双创基地建设项目 | 白城市 | 1.55 | 7.79 |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 扶余市国有粮库粮食仓储物流设施新建项目 | 扶余市 | 10.59 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 珲春综合保税区保税物流分拣中心建设项目 | 珲春市 | 2.30 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 蛟河市职业教育产教融合实训基地建设项目 | 蛟河市 | 1.05 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 靖宇县中医院异地医疗智慧化新建项目 | 靖宇县 | 1.35 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 辽源市新能源汽车配套产业园区基础设施建设项目 | 辽源市 | 12.99 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 延吉高新区产业园区配套设施人才服务中心建设 | 延吉市 | 4.76 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 伊通满族自治县第一人民医院异地新建项目 | 伊通县 | 12.73 | 7.89 |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 农安县康乐乡农业基地建设 | 农安县 | 0.25 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 扶余市城区产业新城建设及道路工程续建 | 扶余市 | 1.39 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 洮北区电力产业园建设项目 | 白城市洮北区 | 6.53 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 敦化市老旧小区改造提升供热系统升级改造 | 敦化市 | 1.80 | |
| 2024年吉林省棚改专项债券（二期）——2024年吉林省政府专 | 敦化市 2023 年城镇老旧小区改造配套设施 | 敦化市 | 1.85 | |
| 2024年吉林省收费公路专项债券（一期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 49.36 | |
| 2024年吉林省收费公路专项债券（二期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 15.41 | |
| 2024年吉林省收费公路专项债券（三期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 16.54 | |
| 2024年吉林省收费公路专项债券（四期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 4.56 | |
| 2024年吉林省收费公路专项债券（五期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 22.02 | |
| 2024年吉林省收费公路专项债券（六期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（七期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（八期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（九期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十一期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十二期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十三期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十四期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十五期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十六期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十七期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十八期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（十九期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十一期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十二期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十三期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十四期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十五期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十六期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十七期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十八期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（二十九期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十一期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十二期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十三期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十四期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十五期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十六期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十七期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十八期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（三十九期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十一期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十二期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十三期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十四期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十五期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十六期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十七期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十八期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（四十九期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | 长春市 | 1.00 | |
| 2024年吉林省收费公路专项债券（五十期）——2024年吉林省政 | 吉林至长春高速公路改扩建工程 | | | |