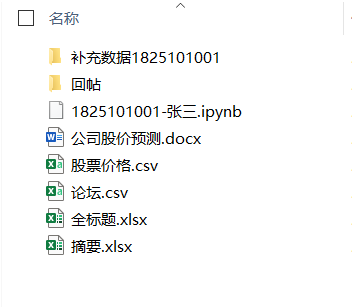
**公司股价预测作业要求**

最终提交：**班级-学号-姓名.ipynb（必须）+ 补充数据 （可选）**

**以 班级-学号-姓名.ipynb的形式（如1班-1825101001-张三） 提交最终作业**

**不得改变数据和代码文件的相对位置；**

**如有补充数据，补充的所有数据文件放置于于 ”补充数据学号” 目录下，数据也需打包提交**



要求：

设计一个预测股票价格的方法，并用实例证明此方法的有效性。

所给的数据，要求全部都要使用，注意

1. 数据需清洗
2. 特征综合使用

可自己额外补充资源或数据

基础得分点

1. **程序可运行**
2. 缺失数据处理
3. 正则提取
4. 时间信息转换
5. 数据拼接
6. 数据标准化
7. 数据可视化
8. 模型构建与应用
9. Markdown 排版
10. 分析逻辑

提供的数据说明：

1. 全标题
   1. 这是股票平台上发布的对各公司的分析文章
   2. 标题：文章的标题
   3. 字段1\_链接\_链接：原文章所在的URL
   4. ABOUT：文章针对的公司，都为缩写形式，多个公司以逗号隔开
   5. TIME：文章发布的时间
   6. AUTHOR：作者
   7. COMMENTS：采集时，文章的被评论次数
2. 摘要
   1. 这是股票平台上发布的对各公司的分析文章的摘要部分，和“全标题”中的内容对应
   2. 标题：文章的标题
   3. 字段2：文章发布的时间
   4. 字段5：文章针对的公司及提及的公司；
      1. About为针对公司，都提取缩写的大写模型，多个公司以逗号隔开
      2. include为提及的其它公司，都提取缩写的大写模型，多个公司以逗号隔开
   5. 字段1：摘要的全文字内容
3. 回帖
   1. 这是网友在各文章下的回复内容
   2. Title：各文章的标题；空标题的，用最靠近的有内容的下方标题
   3. Content：回复的全文字内容
4. 论坛
   1. 这是网友在各公司的论坛页面下，对之进行评论的发帖内容
   2. 字段1：作者
   3. 字段2：发帖日期
   4. 字段3：帖子内容
   5. 字段4\_链接：具体的各公司的页面URL
5. 股票价格
   1. 为各公司工作日股票的价格
   2. PERMNO：公司编号
   3. Date：日期
   4. TICKER：公司简写
   5. COMNAM：公司全写
   6. BIDLO：最低价
   7. ASKHI：最高价
   8. PRC： 收盘价
   9. VOL：成交量
   10. OPENPRC： 开盘价