

机试题目-数据获取与数据处理

机试题目-数据获取与数据处理

概述

题目

资源

数据获取

数据存储

数据处理

要求

补充

概述

爬取指定网站的数据，将获取的数据存入数据库。sqlserver，mysql，sqlite为可选数据库。读取数据库数据，按要求进行数据处理，并输出结果文件。

题目

爬取 天天基金 上的产品数据，需要爬取20个以上产品的基金净值数据。基金净值数据要求为产品存续期的所有净值数据。爬取成功后按定义的数据结构（根据需求与理解自行定义），存入数据库。至此，数据库中存储了至少20个以上的产品基金净值数据。读取数据库中的数据，对净值数据进行处理和输出。

资源

天天基金 <http://fund.eastmoney.com/>

产品页面示例 <http://fund.eastmoney.com/290008.html>

数据获取

从 天天基金 网站获取产品净值数据为主要要求。并不限制在某个页面获取数据，可自行调整和发挥。

数据存储

代码里需要有完整的数据库操作。附件中需要提交数据表格结构定义文件，以及将数据库全量数据的导出到csv文件。

为了保证代码在本地可运行，增加使用文本读写替代数据库读写的操作代码。

数据处理

需要计算各个产品净值数据的评估指标，如年化收益，夏普值，最大回撤等评估指标，并输出文件。

要求

1. 请认真审题。
2. 数据文件字段为 `TradingDay` 与 `Return`，表示交易日与当日收益。
3. 计算每个产品的总收益，年化收益，夏普值，最大回撤。
4. 计算每个产品每年的年收益，年夏普值，年最大回撤。
5. 以上信息需要在一张表输出，行代表产品，列为各项指标。命名需要准确易读。如2018年的夏普可以命名为 `2018_sharpe`，总夏普可以为 `sharpe` 等。
6. 计算每个产品每月收益，并汇总。要求输出表格的列为 `year, month, fund_name, return`。（内容匹配即可，不约束变量命名）
- 7.

补充

1. 请提供完整的可运行代码.py，及对应的书面报告及输出结果文件。代码请加注释说明。
2. 尽可能使用简洁高效的代码。
3. 注意编码，变量名，函数名，类名等命名规范。
4. 开发语言限制为Python。
5. 本题考察范围：爬虫，数据库，数据处理，代码书写习惯，语法基础，编程技巧，金融公式掌握情况。