

Motivation.txt

تمرین پایانی پایتون

مهلت تمرین:

23 دی ماه 1402، ساعت 12 ظهر

(2023 Jan 13)

نمره: 1000

استاد مجتبی بنائی

دوره مهندسی داده فیلاگر





شرح تمرین :

هدف از این تمرین، علاوه بر کار با پایتون، آشنایی با مسایلی است که یک مهندس داده در کارهای روزانه با آنها سروکار دارد. از جمله این کارها، دریافت داده از منابع مختلف و استخراج اطلاعات و نهایتاً ذخیره آنها است. (فرآیند ETL)

برای این تمرین قصد داریم فایل‌های خلاصه داده و ستد روزانه سازمان بورس که در این آدرس قرار دارند را دریافت، پاکسازی، پردازش و ذخیره کنیم :

<http://members.tsetmc.com/tsev2/excel/MarketWatchPlus.aspx?d=1402-10-09>

همانطور که در آدرس بالا مشاهده میکنید، انتهای آدرس حاوی تاریخ روزی است که قرار است اطلاعات آنرا دریافت کنید.

#Filoger

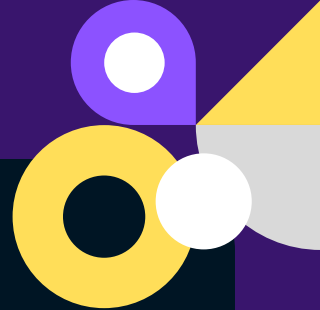




خروجی‌های اصلی تمرین:

-فایل‌های روزانه بورس در یک بازه زمانی با قالب csv که در یک پوشه با نام datalake (دریاچه داده) ذخیره شود.

-مجموعه ای از فایل‌های پایتون برای دریافت و پردازش فایل‌های موجود در دریاچه داده (منظور از دریاچه داده، پوشه ای است که فایل‌ها به صورت خام در آن ذخیره شده اند)



فایل‌های پایتون مورد نیاز:

-برنامه ای بنویسید که در خط فرمان (argparser) بازه زمانی مورد نیاز را گرفته و فایل‌های موجود در آن بازه را در پوشه موقت stage ذخیره کند.

• آیا می‌توانید روزهای پنج‌شنبه و جمعه که بورس تعطیل است را حذف کنید و تنها فایل‌های داد و ستد بورس برای روزهای کاری را دریافت کنید.

-برنامه ای بنویسید که فایل‌های بدون رکورد (مثلاً آنهایی که برای روزهای تعطیل بوده است) را حذف کرده و آنها را به قالب csv تبدیل کند. این فایل هم چند پارامتر خط فرمان دارد که آدرس پوشه stage و اینکه آیا فایل‌های اکسل اولیه را بعد از تبدیل به csv حذف کند یا نه جزء این پارامترها خواهند بود.

-برنامه ای بنویسید که فایل‌های csv را دریافت کرده و چند تحلیل روی آنها انجام دهد (حداقل سه عدد). (انتخاب آنها بر عهده شما: مثلاً ده نماد بورسی با بیشترین حجم معامله در آن بازه، بیشترین افت قیمت، بیشترین رشد قیمت و ...)



ملاحظات :

-دو فایل اول را به صورت مجموعه ای از توابع مختلف بنویسید اما فایل سوم که برای پردازش فایل‌های csv است حتما به صورت کلاس نوشته شود.

-تمامی توابع حاوی بخش try و except باشند و در قسمت except ، توضیحات خطا را در فایل error.log ذخیره کنید.

به جای پرینت کردن مراحل مختلف کار از ماژول logging پایتون استفاده کنید و تمامی پیام‌های برنامه مانند این که مثلا فایل دیتای تاریخ x با موفقیت دانلود شد را به کمک این ماژول و در فایل info.log ذخیره کنید.

موفق باشید