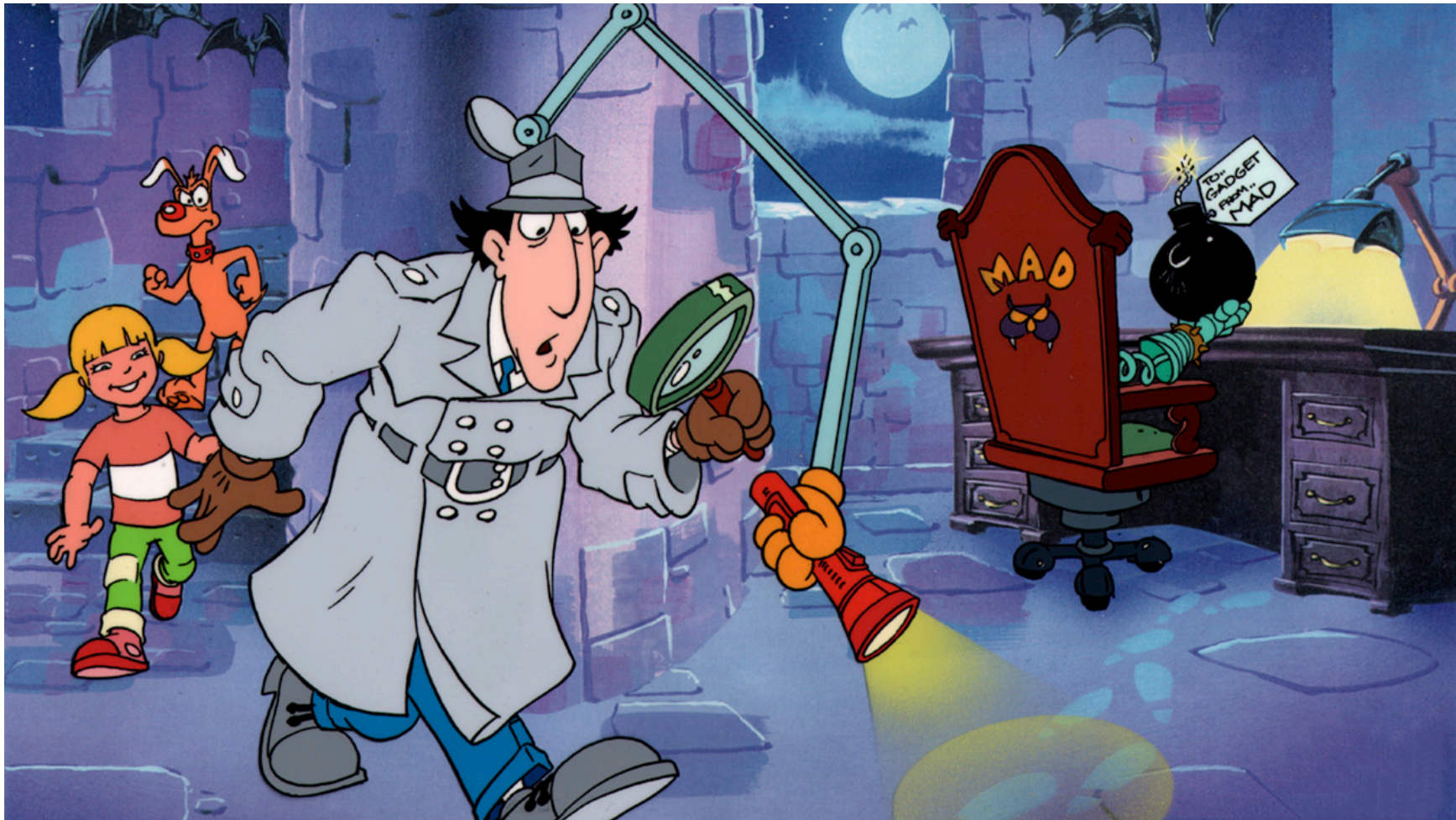


شغل جدید

در هر آشوب و جرمی که در شهر رخ می‌دهد، کارآگاه گجت به دنبال شناسایی عوامل آن می‌گردد. با توجه به این که در یک روز، اتفاقات زیادی در شهر رخ می‌دهد، او نمی‌تواند از همه آن‌ها آگاه شود و طبق اهمیت آن‌ها را اولویت‌بندی کند.



آقای گجت برای بهبود عملکرد خود در ماموریت‌ها، شما را به عنوان دانشمند داده به تیم خود اضافه کرده‌است تا سامانه جامع اخبار (سجا) را ایجاد کنید و برسید به داد آن ناتوان. 😊

در اولین هفته‌ی کاری، شما اقدام به نوشتن برنامه‌ای کردید که تعدادی سایت خبری را هر نیم‌ساعت یک‌بار رصد کرده و اقدام به ذخیره‌سازی اخبار آن‌ها می‌کند.

مجموعه داده

مجموعه داده را می‌توانید از [این لینک](#) دریافت کنید.

بعد از unzip کردن مجموعه داده، پوشه‌هایی را مشاهده می‌کنید که نامشان نشان‌دهنده یک تاریخ است. هرکدام از این پوشه‌ها، شامل پوشه‌هایی با نام منابع خبری هستند که در آن تاریخ رصد شده‌اند و اخبار آن‌ها جمع‌آوری شده است. داخل هر کدام از پوشه‌های مربوط به منابع خبری، فایل‌های html جمع‌آوری شده از آن منبع خبری قرار گرفته‌اند.

```
data
├── 2023-08-05
│   ├── 7sobh
│   │   ├── 0.html
│   │   └── 1.html
│   └── farsnews
│       ├── 0.html
│       └── 1.html
└── 2023-08-04
    ├── varzesh3
    │   ├── 0.html
    │   └── 1.html
    └── snn
```

```
|— 0.html  
|— 1.html
```

هنگامی که یک فایل html را باز می‌کنید، با این ساختار روبه‌رو می‌شوید.

در داخل برچسب head ، شما به آدرس اینترنتی خبر در سایت منبع دسترسی دارید. همچنین در داخل برچسب article ، با رجوع به برچسب h2 می‌توانید به تیتر خبر دسترسی پیدا کنید. اگر خبری دارای عکس باشد، عکس اصلی آن شناسایی شده و در داخل برچسب img قرار گرفته است. در ادامه، متن خبر را در قالب برچسب‌های p آورده‌ایم.

اگر این خبر را از طریق یک مرورگر باز کنید، با تصویر زیر روبه‌رو می‌شوید:

مسیر قیمت طلا در هفته جاری معلوم نیست



به گزارش «کالاخبر» به نقل از ایبسا، تحلیلگران در گفت و گو با کینکوبوز خاطرنشان کردند که با ادامه رشد فشارهای تورمی، بازار طلا شاهد حمایت بنیادین قوی است. با این حال دورنمای فنی طلا منفی است زیرا قیمت این فلز با وجود افت بارز اوراق قرضه و دلار ضعیف تر، موفق نشده است رشدی که در بالای مرز ۱۹۰۰ دلار داشت را حفظ کند.

اگرچه نظر تحلیلگران وال استریت در یک هفته گذشته تغییر فاحشی پیدا کرده است اما سرمایه گذاران غیرحرفه ای همچنان نسبت به افزایش قیمت طلا خوش بین مانده اند.

در نظرسنجی هفته جاری کینکوبوز از تحلیلگران وال استریت ۱۵ نفر شرکت کردند و هر کدام از سه سناریوی بیج رای آوردند که نتیجه غیرمنتظره ای بود.

در این بین، در نظرسنجی آنلاین کینکوبوز ۱۰۵۶ نفر شرکت کردند که از میان آنها ۶۹۵ نفر معادل ۶۶ درصد به افزایش قیمت طلا و ۱۸۱ نفر معادل ۱۷ درصد به کاهش قیمت طلا رای دادند و ۱۸۰ نفر معادل ۱۷ درصد نظری نداشتند.

در نظرسنجی هفته پیش، سرمایه گذاران خرد و تحلیلگران وال استریت نسبت به افزایش قیمت طلا خوش بین بودند اما با بسته شدن قیمت طلا در سطح پایین تر از هفته گذشته، تحلیلگران این خوش بینی را از دست داده اند.

در معاملات روز جمعه بازار آمریکا هر اونس طلا ۱۶ دلار و ۸۰ سنت معادل ۰.۹ درصد کاهش یافت و در ۱۸۷۹ دلار و ۶۰ سنت بسته شد و برای کل هفته حدود ۰.۷ درصد کاهش نشان داد.

تحلیلگران می گویند از نظر بنیادین فروش طلا دشوار است چرا که تورم ماه گذشته شاهد افزایش سالانه پنج درصد بود که بزرگترین افزایش از آگوست سال ۲۰۰۸ به شمار می رود. رشد تورم به همراه نرخ های بهره پایین تر به معنای آن است که بارزده واقعی منفی تر شده است.

بسیاری از تحلیلگران بر این باورند که هر چند قیمت طلا در حال تحکیم است اما روند این فلز افزایشی مانده زیرا قیمت طلا بالای سطح ۱۸۵۵ دلار را حفظ کرده است.

اوله هانسن، مدیر استراتژی کالا در ساکسو بانک اظهار کرد: اگرچه از نظر فنی نسبت به طلا نظر منفی دارد اما دلیلی نمی بیند که فروشنده باشد. بازار اندکی از نیروی صعودی خود را از دست داده و باید محرک جدیدی برای روند صعودی پیدا کند.

دارین نیوسام، رئیس شرکت دارین نیوسام آنالیز به کینکوبوز گفت: وی نظر منفی نسبت به قیمت طلا در کوتاه مدت دارد و خاطرنشان کرد منتظر اصلاح بیشتر قیمت طلاست.

فیلیپ استرایبل، استراتژیست ارشد بازار در شرکت بلو لاین فیوجرز اظهار کرد: انتظار دارد قیمت طلا در کوتاه مدت پایین تر برود با این حال این کاهش به حدی نخواهد بود که وی نسبت به طلا بدبین شود. وی ترجیح می دهد منتظر بماند طلا تا سطح ۱۸۵۰ دلار نزول کند تا دوباره خرید را شروع کند.

مارک چندلر، مدیرکل شرکت بانوک برن گلوبال فارکس اظهار کرد: وی نسبت به قیمت طلا بدبین است زیرا طلا در سطح ۱۹۰۰ دلار با مقاومت نیرومندی روبرو شده است. وی پیش بینی کرد بارزده اوراق قرضه در آستانه نشست سیاست پولی بانک مرکزی آمریکا صعود کند.

صورت مسئله

در ابتدای کار، شما می‌خواهید با بررسی اولیه به اکتشاف در مجموعه داده بپردازید و به سوالات زیر پاسخ دهید (مثال خروجی هر سوال، زیر آن آمده است).

۱. چند منبع خبری یکتا در مجموعه داده وجود دارند؟

10

۲. کدام منبع خبری، در کل مجموعه داده بیشترین تعداد فایل‌های html را دارد؟

qnews

۳. تمام فایل‌های html مربوط به varzesh3 را در نظر بگیرید. به ترتیب تعداد تکرار کلمات والیبال، کشتی و فوتبال را در تمام تگ‌های p به دست آورید.

توجه ▼

- دقیقاً خود این کلمات مد نظر هستند و برای مثال کلمه فوتبالش یا کشتی‌گیر نباید شمارش شود.
- متن‌ها دارای علائم نگارشی هستند که باید به آنها توجه کنید.

در خروجی زیر، مقدار 100 مربوط به والیبال، مقدار 200 مربوط به کشتی و مقدار 300 مربوط به فوتبال است.

100,200,300

۴. در بین تمام خبرهایی که در تاریخ 2023-08-01 جمع‌آوری شده‌اند، بیشترین کلمه‌ای که در تگ h2 تکرار شده است را به دست آورید. (می‌توانید از آرایه stopwords قرار داده شده در فایل اولیه استفاده کنید)

قیمت

این مرحله در مجموع، شامل ۴ سوال (هر سوال، ۲۵ امتیاز) است. در صورتی که به تمامی سوالات پاسخ صحیح دهید، بایستی ۱۰۰ امتیاز از این مرحله کسب کنید.

خروجی

برای ارسال پاسخ، جواب هر سوال را مشابه مثال، در یک سطر از فایل text بنویسید. در نهایت یک فایل با فرمت txt. چهار خطی با نام output.txt برای ما بارگذاری کنید. در صورتی که جواب سوالی را نمی‌دانید در سطر مربوطه ۱- وارد کنید.

▼ هشدار

فراموش نکنید که قبل از پایان زمان مسابقه، باید تمامی کدهای این مسابقه را از قسمت بارگذاری گُذ برای ما ارسال کنید. در غیر این صورت، شما از این مسابقه، امتیازی کسب نمی‌کنید.

توجه داشته باشید که اگر از jupyter notebook استفاده می‌کنید بایستی همانند توضیحات قسمت بارگذاری گُذ، خروجی py. را دریافت و برای ارسال در نظر بگیرید. ارسال فایل‌های jupyter همانند ipynb. مورد قبول نیستند.

نمونه خروجی ۱

Copy Plain text

10

qnews

100,200,300

قیمت

▼ توضیحات

در خط /ام جواب سوال /ام آمده است.

نمونه خروجی ۲

10

-1

-1

قیمت

▼ توضیحات

چون پاسخ سوال دوم و سوم را نمی دانستید، به جای آن عدد -۱ نوشته اید!