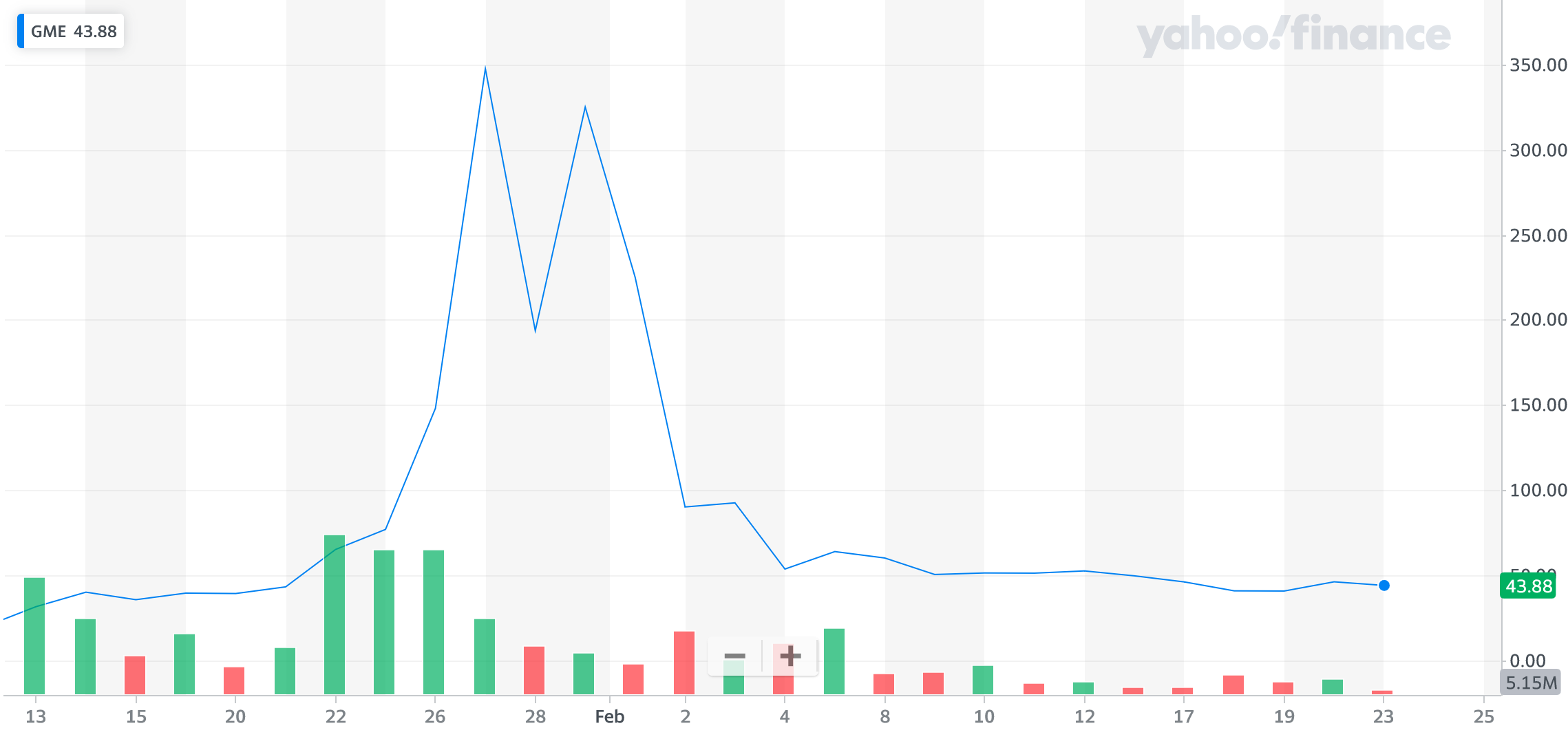
# Business Context

(we are a) Marketing / Content Creator advisory firm

(ourselves or client) really wants to get in on the GME stock hype

the stock is now returning to near-normal levels. Wondering if it's too late to begin talking about this, and if other topics should be pursued



Goal: Pull and store data from the New York Times to use for analyses to find what events are considered current or popular news

About the Data

* Multiple different APIs available, all serving different services
* Depth of the data between responses varies, both between APIs and between response objects from a single API

(e.g. of different responses here?)

## Approaches

Approach 1: Capture raw copies of responses to a database (e.g. an OLTP system)

Goal/Priority: Simply have a copy saved for easy processing or to have archived

Databases will serve as a simple copy of data for local use

Use MongoDB

Structure is adapted for flexible for fast processing of the varied responses available between APIs and responses

Approach 2: Deliver cleaned copies of responses directly to a database, ready for analysis (e.g. an OLAP system)

Goal/Priority: Immediate deployment of analysis ready data

Database will serve as a cleaned data source for immediate processing of specific analyses (e.g. finding counts of popular search terms, finding counts of articles

Send to PostgreSQL database

Approach 3:

Combination of approaches 1 and 2

Goal/Priority: Responsiveness to new data and provides template data structures for frequently queried analyses

save local copies to OTLP system (which uses MongoDB) to ensure copies are available for multiple different ways to process

From MongoDB, make queries from local copies, and send filtered data to OTAP system (which uses PostgreSQL)

## ETL Focus on Approach 1

### General

Capture raw copies of responses to a database (e.g. an OLTP system)

### Prep

Create MongoDB database and collection

### Extract

create loop if necessary by the context of the API's purpose

call to API

retrieve response

(E.g.'s here)

### Transform

minor modifications to clean of metadata and get main objects, building a new dictionary that is "the data set that makes the most sense as a single unit"

(E.g.'s here)

### Load

post dictionary to mongo db as a collection

(E.g.'s here)

### Other

#### Execution Frequency:

(Weekly?)

## ETL Focus on Approach 2

### General

Retrieve raw API responses from the New York Times API and immediately prepare data for a series of analysis

### Prep

Create PostgreSQL db and table

### Extract

make queries to MongoDB via pymongo

### Transform

pull out specific values to variables

(E.g.'s here)

change data types or append to lists/dictionaries/pd dataframes

### Load

form pandas dataframe and export to csv, or post directly to PostgreSQL

### Other

#### Execution Frequency:

(whenever the analysis is requested?)

## ETL Focus on Approach 3

### General

Combination of 1 and 2

Part1:

ETL exactly as approach 1

Part2:

requests using mongodb

### Prep

Create SQL table

### Extract

call to API

retrieve response

### Transform

Pull out specific data

### Load

post directly to PostgreSQL using SQL alchemy

### Other

#### Execution Frequency:

(weekly?)

# Potential Analysis

## Questions to ask

### What are the actual top topics being talked about right now?

#### How to answer

find count of top keywords founds in articles

### Is the Gamestop stock still relevant in today's discussions?

#### How to answer

find count of articles by date by keyword

### What are some other sources that are talking about Gamestop right now?

#### How to answer