# Vector Databases can help with Anamoly detection

For this we nees a dataset with anamolies why not use a dataset that is live auto updating so that constantly anamolies take place which can be auto detected and presented in a readable format for the user to read and apply to the real world.

For this tutorial, I will be using [1] .

## Scraping to create a dataset

Let’s see about all the content that is availaible

import requests

from bs4 import BeautifulSoup

import pandas as pd

# URL for Rediff Money news

url = "https://money.rediff.com/news"

# Fetch HTML content using requests

response = requests.get(url)

if response.status\_code == 200:

    # Parse HTML content using BeautifulSoup

    soup = BeautifulSoup(response.content, "html.parser")

soup

You will get the HTML for all the data available.

Let’s convert this to a proper scraped dataset

import requests

from bs4 import BeautifulSoup

import pandas as pd

# URL for Rediff Money news

url = "https://money.rediff.com/news"

# Fetch HTML content using requests

response = requests.get(url)

if response.status\_code == 200:

    # Parse HTML content using BeautifulSoup

    soup = BeautifulSoup(response.content, "html.parser")

    # Find all news items

    news\_items = soup.find\_all("div", class\_="rtnews\_row\_more")

    # Extract relevant information from each news item

    news\_data = []

    for item in news\_items:

        title = item.find("p").text.strip()

        link = item.find("a")["href"]

        summary = item.find("div").text.strip()

        published = item.find("span", class\_="timeago").text.strip()

        news\_data.append({

            "title": title,

            "summary": summary,

            "link": link,

            "published": published

        })

    # Create pandas DataFrame

    df = pd.DataFrame(news\_data)

    # Save DataFrame to CSV

    df.to\_csv("Dataset/financial\_news.csv", index=False)

    print("Financial news scraped and saved to financial\_news.csv")

else:

    print("Failed to fetch the webpage")

Financial news scraped and saved to financial\_news.csv

Let’s view it

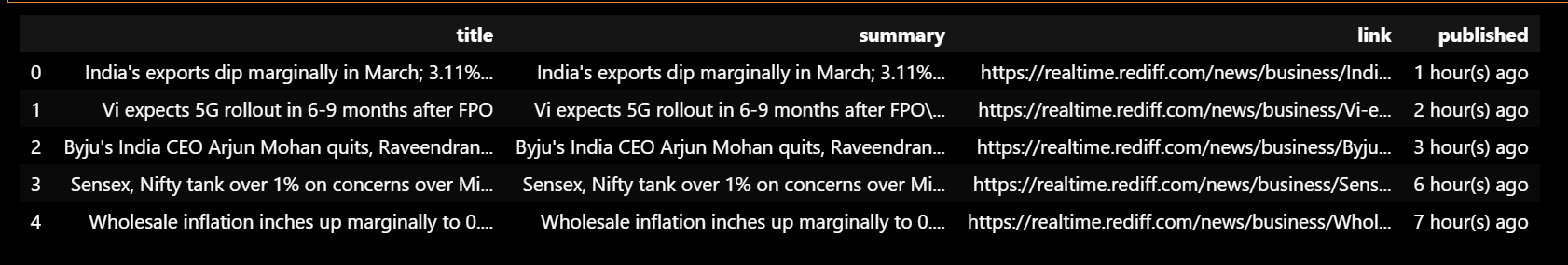
import pandas as pd

# Read the CSV file into a DataFrame

df = pd.read\_csv("Dataset/financial\_news.csv")

# Display the first few rows of the DataFrame

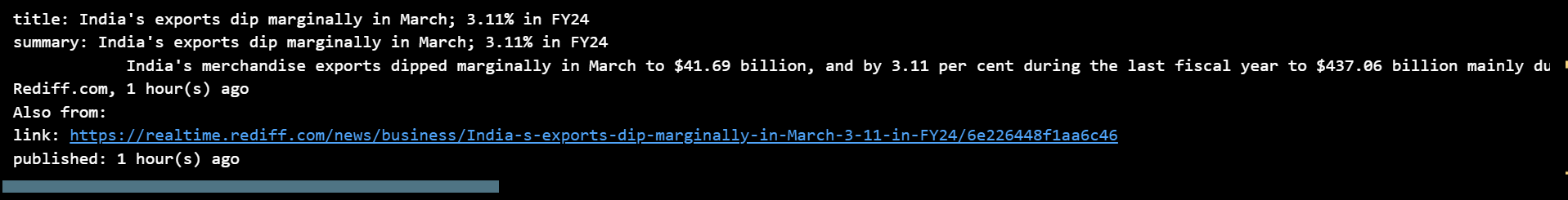
df.head()



Instead of using the pre made dataset using a Live dataset is much better because this will allow the user to check out the concerned stock By Vector database anamoly detection. The user can take BUY and SELL decisions on that stock.

for column in df.columns:

    print(f"{column}: {df[column][0]}")



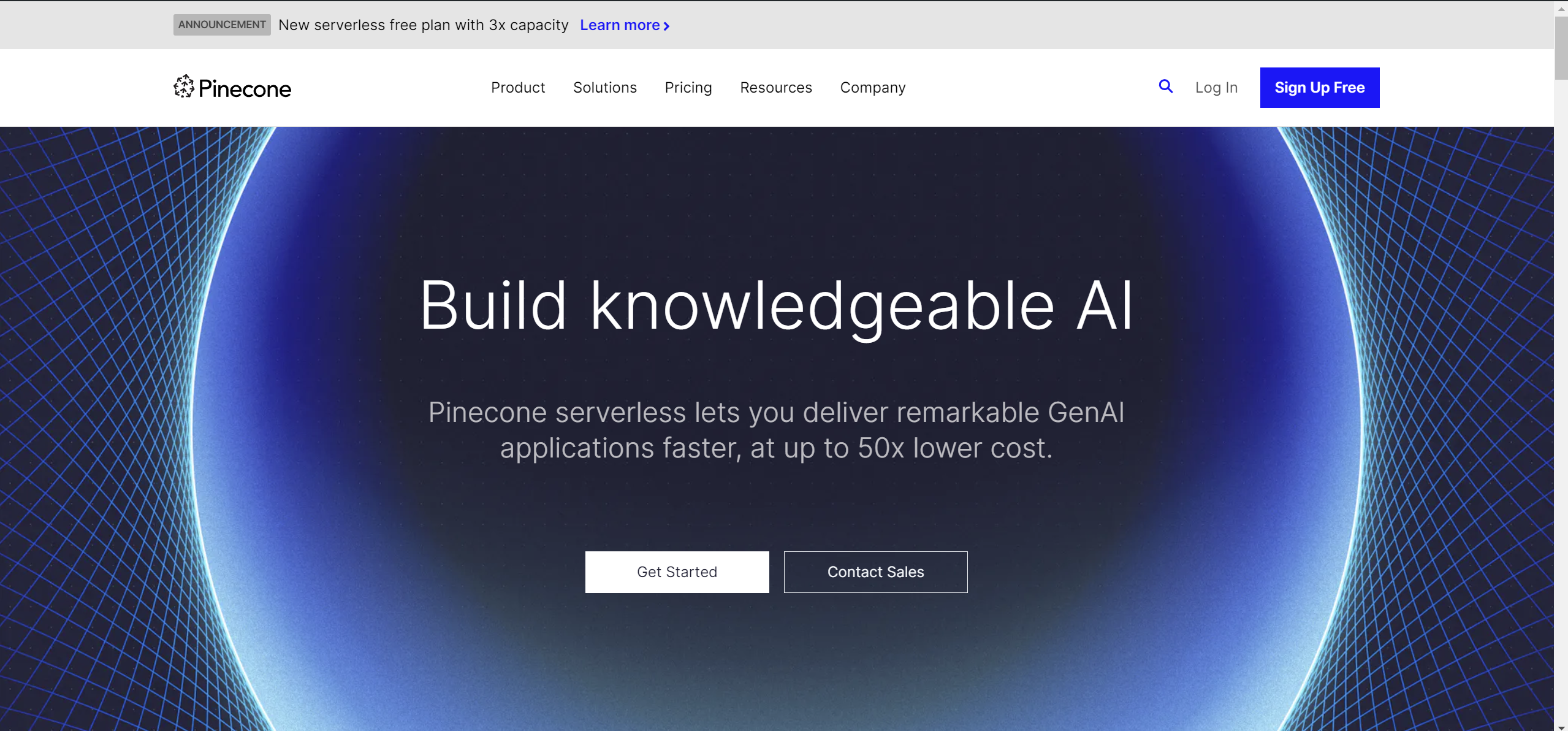
Feeding the dataset to Vector Store

In the last project [2], I used Qdrant which is a locally supported Vector Store for RAG

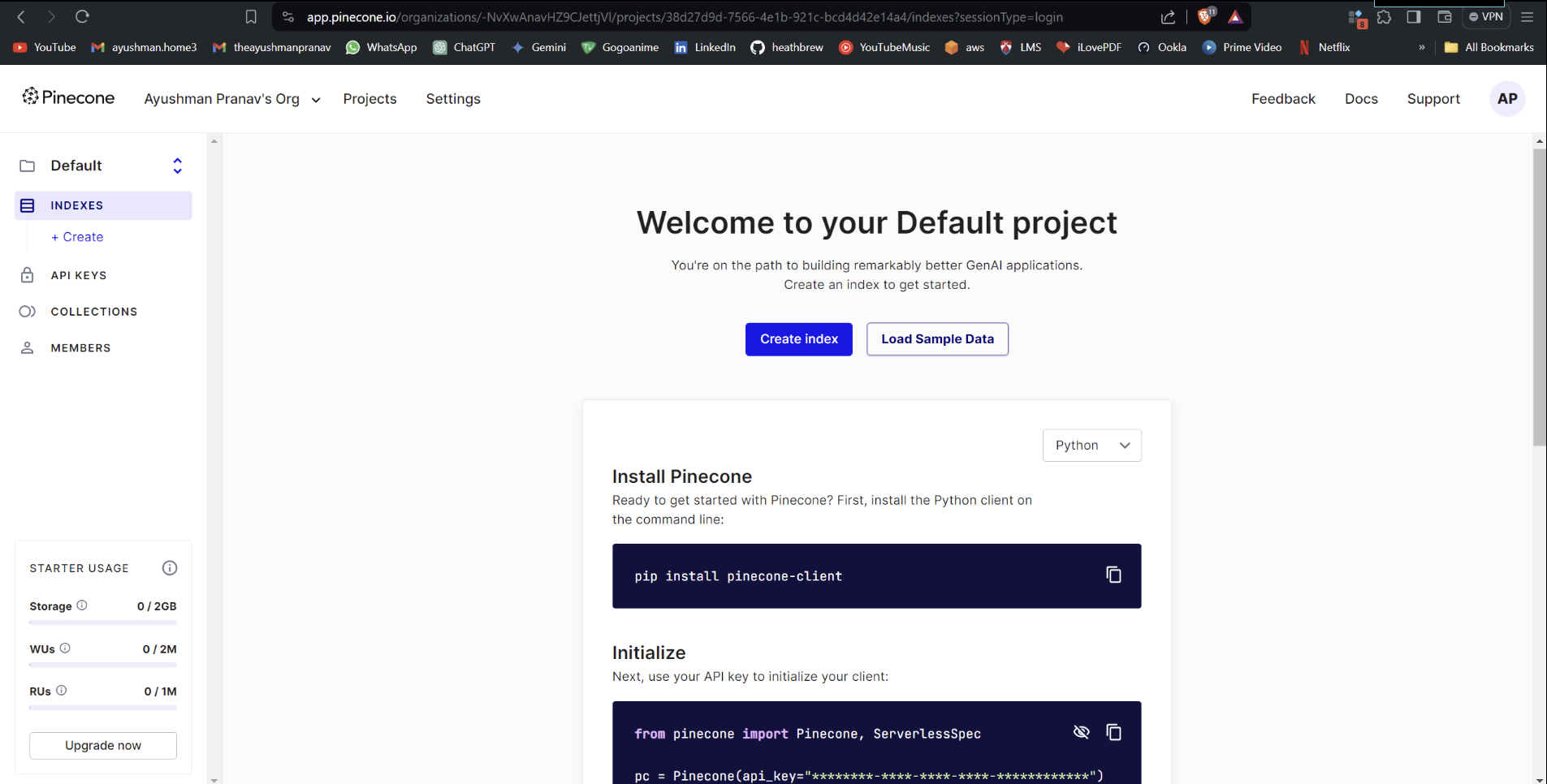
<https://medium.com/devops-dev/steps-to-monitoring-dspy-qdrant-powered-rag-with-prometheus-or-grafana-b642335cbd50>

For this project, I will use PineCone which is an AWS supported Vector store that you can directly In any Live Project

<https://www.pinecone.io/>



Login and create the account



Go to api keys to create your own api key

