

```
Al_Powered_Financial_Analysis.ipynb
File Edit View Insert Runtime Tools Help
Q Commands + Code + Test
StructField("url", StringType(), True),
StructField("title", StringType(), True),
StructField("date", StringType(), True),
StructField("content", StringType(), True),
})

[12] article_dataframe = session.createDataFrame(filtered_articles_rdd, schema=article_structure)

article_dataframe.show(truncate=100)

article_dataframe.write.mode("overwrite").json("output/realtime_financial_data")

+-----+-----+-----+
| url | title | date |
+-----+-----+-----+
| https://www.cnn.com/2025/04/16/money/trump-to-avoid-in-the-market-fast-money-trader-tim-seymour-wants- | Four money traps to avoid in a volatile market, according to "fast money" trader Tim Seymour | 2025-04-16 00:00:00 | Tim Seymour wants |
| https://www.cnn.com/2025/04/17/stocks/international-equities-heres-what-experts-are-saying- | Should investors dump U.S. stocks for international equities? Here's what experts are saying | 2025-04-17 00:00:00 | watch now |
| https://www.cnn.com/2025/04/17/stocks/making-the-biggest-moves-promarket-bag-bag-jaj.html | Stocks making the biggest moves promarket: Bank of America, Boeing, Johnson & Johnson and more | 2025-04-15 00:00:00 | Check out the companies making headline |
| https://www.cnn.com/2025/04/16/hertz-surges-after-bill-ackman-takes-big-stake-in-the-rental-car- | Hertz shares surge more than 50% after Bill Ackman takes big stake in rental car firm | 2025-04-16 00:00:00 | Shares of Hertz surged 56% on Wednesday |
| https://www.cnn.com/2025/04/16/stocks/making-the-biggest-moves-promarket-us- | Stocks making the biggest moves promarket: United Airlines, Nvidia, Tesla, Travelers and more | 2025-04-16 00:00:00 | Check out the companies making headline |
| https://www.cnn.com/2025/04/16/capital-one-and-discover-merger-approved-by-federal-reserve-board- | Capital One and Discover merger approved by Federal Reserve | 2025-04-16 00:00:00 | Capital One Financial's application to |
| https://www.cnn.com/2025/04/17/american-express-cardholders-still-seeing-dislike-trump-tariffs- | American Express' wealthy cardholders are mostly untouched by tariff jitters | 2025-04-17 00:00:00 | American Express' affluent cardholder |
| https://www.cnn.com/2025/04/17/stocks/making-the-biggest-moves-promarket-htr-unh-illy.html | Stocks making the biggest moves promarket: Hertz, UnitedHealth, Eli Lilly and more | 2025-04-17 00:00:00 | Check out the companies making headline |
| https://www.cnn.com/2025/04/19/trump-approval-rating-on-economy-drops-to-lowest-of-his-presiden- | Trump's approval rating on the economy drops to lowest of his presidency | 2025-04-19 00:00:00 | U.S. President Donald Trump speaks as |
| https://www.cnn.com/2025/04/19/cash-are-fool-safes-when-stocks-slide-but-it-has-its- | Cash may feel safe when stocks slide, but it has risks | 2025-04-19 00:00:00 | Traders work on the floor of the New Y |
| https://www.cnn.com/2025/04/20/close-stocks-that-could-survive-delisting-tariff-worries.html | Bernstein predicts which Chinese stocks can survive delisting and tariff worries | 2025-04-20 00:00:00 | Bernstein is looking at Chinese intern |
| https://www.cnn.com/2025/04/21/china-to-retaliate-against-us-after-week-long-trade-tariffs- | China vows retaliation against countries that follow U.S. calls to isolate Beijing | 2025-04-21 00:00:00 | China on Monday warned it w |
| https://www.cnn.com/2025/04/17/ubs-losses-crown-as-continental-europe-s-most-valuable-bank-to-sa- | UBS loses crown as continental Europe's most valuable bank to Santander amid U.S. tariffs | 2025-04-17 00:00:00 | Spanish leader Rocco Santander has ecl |
| https://www.cnn.com/2025/04/16/heres-why-retirees-shouldnt-fully-ditch-stocks.html | Here's why retirees shouldn't fully ditch stocks | 2025-04-16 00:00:00 | Lorden |
| https://www.cnn.com/2025/04/16/nyse-residents-systems-are-handling-record-volumes-more-efficient- | NYSE system are working normally, handling record volume more efficiently than during Cov... | 2025-04-16 00:00:00 | the great American entrepreneur Henry J |
| https://www.cnn.com/2025/04/16/jpmorgan-chase-infinitely-money-glitch-bank- | JPMorgan Chase sues more customers who allegedly stole cash in 'infinite money glitch' | 2025-04-16 00:00:00 | A person uses an ATM at a Chase bank i |
| https://www.cnn.com/2025/04/18/where-made-in-china-2025-missed-the-mark.html | Where 'Made in China 2025' missed the mark | 2025-04-18 00:00:00 | Smart robotic arm work on the product |
| https://www.cnn.com/2025/04/17/trump-again-calls-for-fed-to-cut-rates-says-powell-trendation-c- | Trump again calls for Fed to cut rates, says Powell's 'trendation cannot come fast enough' | 2025-04-17 00:00:00 | President Donald Trump on Thursday aga |
| https://www.cnn.com/2025/04/19/investor-protection-during-market-volatility-through-tactical-fu- | This fund is designed to help investors withstand wild market swings | 2025-04-19 00:00:00 | Katie Stockton thinks she has a viable |
+-----+-----+-----+
only showing top 20 rows
```

```
Al_Powered_Financial_Analysis.ipynb
File Edit View Insert Runtime Tools Help
Q Commands + Code + Test
reddit_df.show(truncate=True)

[17] reddit_df.collect().write \
    .mode("overwrite") \
    .format("json") \
    .option("compression", "none") \
    .save(data_dir)

print(f"✓ Stored {reddit_df.count()} posts at {data_dir}")
else:
    print(f"✗ No Reddit data retrieved.")

reddit_session.stop()

WARNING:praw! It appears that you are using praw in an asynchronous environment.
It is strongly recommended to use Async PRAW: https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.
See https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.

WARNING:praw! It appears that you are using praw in an asynchronous environment.
It is strongly recommended to use Async PRAW: https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.
See https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.

WARNING:praw! It appears that you are using praw in an asynchronous environment.
It is strongly recommended to use Async PRAW: https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.
See https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.

WARNING:praw! It appears that you are using praw in an asynchronous environment.
It is strongly recommended to use Async PRAW: https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.
See https://praw.readthedocs.io/en/latest/getting-started/multiple_instances.html#discord-bots-and-asynchronous-environments for more info.

+-----+-----+-----+-----+
| Title | URL | Upvotes | Comments | Count | Post Time |
+-----+-----+-----+-----+-----+
| Calls on Q&A | https://l.reddit.it... | 1 | 0 | 2025-04-21 07:29:17 |
| Selling my feet | https://l.reddit.it... | 8 | 0 | 2025-04-21 07:13:01 |
| [Expected Interest... | https://l.reddit.it... | 20 | 24 | 2025-04-21 06:21:39 |
| Buying TSLA calls... | https://www.reddit... | 228 | 92 | 2025-04-21 06:01:10 |
| New | https://www.reddit... | 17 | 17 | 2025-04-21 00:10:22 |
+-----+-----+-----+-----+
only showing top 5 rows

✓ Stored 500 posts at /content/reddit_data_raw
```

```
AI Powered Financial Analysis.ipynb
File Edit View Insert Runtime Tools Help
Q Commands + Code + Text
df_without_changing = ingest_file(input_path, sqlContext)

print("Sample of raw data:")
df_without_changing.show(truncate=False)
print("Schema of raw data:")
df_without_changing.printSchema()

Successfully ingested SEC filings.csv. Total rows: 2622
Sample of raw data:

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| TickerName | Sector | Asset Class | Market Value | Weight (%) | National Value | Quantity | Price | Location | Exchange | Currency | FX Rate | Market Currency | Accrual Date |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| AAPL | APPLE INC | Information Technology | Equity | 599,365,151.11 | 5.16 | 599,365,151.11 | 4,385,127.00 | 129.93 | United States | NASDAQ | USD | 1 | USD | - |
| PGFI | MICROSOFT CORP | Information Technology | Equity | 513,917,732.62 | 4.74 | 513,917,732.62 | 2,142,031.00 | 239.82 | United States | NASDAQ | USD | 1 | USD | - |
| AMZN | AMAZON COM INC | Consumer Discretionary | Equity | 213,623,596.00 | 1.97 | 213,623,596.00 | 2,545,510.00 | 84 | United States | NASDAQ | USD | 1 | USD | - |
| UNH | UNITEDHEALTH GROUP INC | Health Care | Equity | 189,601,687.00 | 1.47 | 189,601,687.00 | 116,064.00 | 1628.9 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| GOOG | ALPHABET INC CLASS A | Communication | Equity | 151,990,826.75 | 1.4 | 151,990,826.75 | 1,722,725.00 | 88.23 | United States | NASDAQ | USD | 1 | USD | - |
| JNJ | JOHNSON & JOHNSON | Health Care | Equity | 142,628,859.84 | 1.31 | 142,628,859.84 | 267,888.00 | 536.18 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| XOM | EXXON MOBIL CORP | Energy | Equity | 139,597,878.61 | 1.25 | 139,597,878.61 | 2,277,297.00 | 61.23 | United States | NASDAQ | USD | 1 | USD | - |
| JPM | JPMORGAN CHASE & CO | Financials | Equity | 133,147,817.70 | 1.23 | 133,147,817.70 | 253,736.00 | 176.05 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| KO | KOCH INDUSTRIES INC | Energy | Equity | 130,179,148.40 | 1.2 | 130,179,148.40 | 1,180,228.00 | 110.3 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| DIS | DISNEY | Entertainment | Equity | 112,822,176.30 | 1.01 | 112,822,176.30 | 835,363.00 | 134.1 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| PG | PROCTER & GAMBLE | Consumer Staples | Equity | 102,458,349.00 | 0.95 | 102,458,349.00 | 676,825.00 | 151.56 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| INFA | INVELOS CORP | Information Technology | Equity | 100,792,173.44 | 0.93 | 100,792,173.44 | 680,696.00 | 146.14 | United States | NASDAQ | USD | 1 | USD | - |
| COW | COWI CORP | Energy | Equity | 99,412,895.29 | 0.92 | 99,412,895.29 | 253,891.00 | 179.49 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| V | VISA INC CLASS A | Information Technology | Equity | 97,955,931.36 | 0.9 | 97,955,931.36 | 1471,486.00 | 287.76 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| HD | HOME DEPOT INC | Consumer Discretionary | Equity | 93,465,816.74 | 0.86 | 93,465,816.74 | 225,969.00 | 315.86 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| TSLA | TESLA INC | Consumer Discretionary | Equity | 91,176,481.82 | 0.84 | 91,176,481.82 | 749,109.00 | 123.18 | United States | NASDAQ | USD | 1 | USD | - |
| LILY | ELLI LILLY | Health Care | Equity | 88,285,240.48 | 0.82 | 88,285,240.48 | 124,322.00 | 365.84 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| INR | INGENUITY INC CLASS A | Information Technology | Equity | 84,738,671.43 | 0.78 | 84,738,671.43 | 243,691.00 | 347.73 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| PFE | PFIZER INC | Health Care | Equity | 82,940,884.52 | 0.77 | 82,940,884.52 | 1,618,675.00 | 51.24 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
| ABBV | ABBVIE INC | Health Care | Equity | 81,797,931.84 | 0.76 | 81,797,931.84 | 158,144.00 | 181.61 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

only showing top 20 rows

Schema of raw data:
root
-- Tickers: string (nullable = true)
-- Name: string (nullable = true)
-- Sector: string (nullable = true)
-- Asset Class: string (nullable = true)
-- Market Value: string (nullable = true)
-- Weight (%): string (nullable = true)
-- National Value: string (nullable = true)
-- Quantity: string (nullable = true)
-- Price: string (nullable = true)
-- Location: string (nullable = true)
-- Exchange: string (nullable = true)
-- Currency: string (nullable = true)
-- FX Rate: string (nullable = true)
-- Market Currency: string (nullable = true)
-- Accrual Date: string (nullable = true)
```

```
AI Powered Financial Analysis.ipynb
File Edit View Insert Runtime Tools Help
Q Commands + Code + Text
cleaned_df = df_without_changing.select(
    col("TickerName"),
    col("Sector"),
    col("Asset Class"),
    col("Market Value"),
    col("Weight (%)" ),
    col("National Value"),
    col("Quantity"),
    col("Price"),
    col("Location"),
    col("Exchange"),
    col("Currency"),
    col("FX Rate"),
    col("Market Currency"),
    col("Accrual Date"),
    col("Market Value"),
    col("Weight (%)")
)

cleaned_df.printSchema()

Data cleaned. Total rows after cleaning: 2622
Sample of cleaned data:

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| TickerName | Sector | Asset Class | Market Value | Weight (%) | National Value | Quantity | Price | Location | Exchange | Currency | FX Rate | Market Currency | Accrual Date | Market Value | Weight (%) |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| L | LUMBI INC 1250 vesting prv | Health Care | Equity | 0.02 | 0 | 0.02 | null | 0.0 | United States | NO MARKET (E.G. UNLISTED) | USD | 1 | USD | - | 0.02 | 0.0 |
| a | agilent technologies inc | Health Care | Equity | 12,728,488.75 | 0.12 | 12,728,488.75 | null | 149.65 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.12 |
| ia | ilico corp | Materials | Equity | 2,398,992.44 | 0.02 | 2,398,992.44 | null | 45.47 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.02 |
| aal | american airlines group inc | Industrials | Equity | 2,398,638.80 | 0.02 | 2,398,638.80 | null | 12.72 | United States | NASDAQ | USD | 1 | USD | - | null | 0.02 |
| aam | aerams company inc | Consumer Discretionary | Equity | 117,779.20 | 0 | 117,779.20 | null | 11.95 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.0 |
| aam | aam inc | Industrials | Equity | 869,305.04 | 0.01 | 869,305.04 | null | 75.32 | United States | NASDAQ | USD | 1 | USD | - | null | 0.01 |
| aap | advance auto parts inc | Consumer Discretionary | Equity | 2,611,105.77 | 0.02 | 2,611,105.77 | null | 147.49 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.02 |
| aapl | apple inc | Information Technology | Equity | 599,365,151.11 | 5.16 | 599,365,151.11 | null | 129.93 | United States | NASDAQ | USD | 1 | USD | - | null | 5.16 |
| ael | american assets trust reit inc | Real Estate | Equity | 355,815.50 | 0 | 355,815.50 | null | 26.5 | United States | New York Stock Exchange Inc. | USD | 1 | USD | Jan 01, 1970 | null | 0.0 |
| aaw | atlas air worldwide holdings inc | Industrials | Equity | 866,275.20 | 0.01 | 866,275.20 | null | 188.8 | United States | NASDAQ | USD | 1 | USD | - | null | 0.01 |
| aebv | aetna inc | Health Care | Equity | 181,797,931.84 | 0.76 | 181,797,931.84 | null | 165.71 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.76 |
| abc | americanairways corp | Health Care | Equity | 7,540,124.45 | 0.07 | 7,540,124.45 | null | 165.71 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.07 |
| abcb | ameris bancorp | Financials | Equity | 769,419.08 | 0.01 | 769,419.08 | null | 47.14 | United States | NASDAQ | USD | 1 | USD | - | null | 0.01 |
| abcl | abbott laboratories | Health Care | Equity | 839,362.20 | 0.01 | 839,362.20 | null | 165.71 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.01 |
| abg | abbey automotive group inc | Consumer Discretionary | Equity | 1,152,577.50 | 0.01 | 1,152,577.50 | null | 179.25 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.01 |
| abh | abn industries inc | Industrials | Equity | 189,362.20 | 0.01 | 189,362.20 | null | 44.42 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.01 |
| abib | abn inc class a | Consumer Discretionary | Equity | 9,181,811.50 | 0.08 | 9,181,811.50 | null | 185.5 | United States | NASDAQ | USD | 1 | USD | - | null | 0.08 |
| abj | arbor realty trust reit inc | Financials | Equity | 593,365.34 | 0.01 | 593,365.34 | null | 13.19 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.01 |
| abt | abbott laboratories | Health Care | Equity | 53,916,551.52 | 0.5 | 53,916,551.52 | null | 169.79 | United States | New York Stock Exchange Inc. | USD | 1 | USD | - | null | 0.5 |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

only showing top 20 rows

Schema of cleaned data:
root
-- Tickers: string (nullable = true)
-- Name: string (nullable = true)
-- Sector: string (nullable = true)
-- Asset Class: string (nullable = true)
-- Market Value: string (nullable = true)
-- Weight (%): string (nullable = true)
-- National Value: string (nullable = true)
-- Quantity: integer (nullable = true)
-- Price: float (nullable = true)
-- Location: string (nullable = true)
-- Exchange: string (nullable = true)
-- Currency: string (nullable = true)
-- FX Rate: string (nullable = true)
-- Market Currency: string (nullable = true)
-- Accrual Date: string (nullable = true)
-- Market Value: float (nullable = true)
-- Weight (%): float (nullable = true)
```

```
AI Powered Financial Analysis.ipynb
File Edit View Insert Runtime Tools Help
Q Commands + Code + Text
[24] print("Sample of metadata")
extracted_metadata.show(truncate=False)
print("Schema of metadata")
extracted_metadata.printSchema()

Metadata extracted successfully.
Sample of metadata:
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| ticker | company_name | sector | market_value | weight_percent | shares_quantity | share_price | country | stock_exchange | accrual_date | record_id |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|-| comiab inc 1250 vesting prvt | health care | 0.02 | 0.0 | null | 0.0 | United States | NO MARKET (E.G. UNLISTED) | - | null | |
|-| comiab inc 1500 vesting prvt | health care | 0.02 | 0.0 | null | 0.0 | United States | NO MARKET (E.G. UNLISTED) | - | null |
|-| agilent technologies inc | health care | null | 0.12 | null | 149.65 | United States | New York Stock Exchange Inc. | - | null |
|-| alisa corp | materials | null | 0.02 | null | 45.47 | United States | New York Stock Exchange Inc. | - | null |
|-| american airlines group inc | industrials | null | 0.02 | null | 12.72 | United States | NASDAQ | - | null |
|-| aerom company inc | consumer discretionary | null | 0.0 | null | 13.95 | United States | New York Stock Exchange Inc. | - | null |
|-| ason inc | industrials | null | 0.01 | null | 75.32 | United States | NASDAQ | - | null |
|-| aspf | advance auto parts inc | consumer discretionary | null | 0.02 | null | 147.63 | United States | New York Stock Exchange Inc. | - | null |
|-| apple inc | information technology | null | 5.16 | null | 129.93 | United States | NASDAQ | - | null |
|-| asat | american assets trust reit inc | real estate | null | 0.0 | null | 26.5 | United States | New York Stock Exchange Inc. | Jan 01, 1970 | null |
|-| asaw | atlas air worldwide holdings inc | industrials | null | 0.01 | null | 180.8 | United States | NASDAQ | - | null |
|-| abbv | abbvie inc | health care | null | 0.76 | null | 161.61 | United States | New York Stock Exchange Inc. | - | null |
|-| abc | americansourceberg corp | health care | null | 0.07 | null | 165.71 | United States | New York Stock Exchange Inc. | - | null |
|-| abcb | ameris bancorp | financials | null | 0.01 | null | 147.14 | United States | NASDAQ | - | null |
|-| abcl | abcellera biologics inc | health care | null | 0.01 | null | 18.13 | Canada | NASDAQ | - | null |
|-| abg | ashbury automotive group inc | consumer discretionary | null | 0.01 | null | 179.25 | United States | New York Stock Exchange Inc. | - | null |
|-| abe | abe industries inc | industrials | null | 0.01 | null | 44.42 | United States | New York Stock Exchange Inc. | - | null |
|-| absh | airbnb inc class a | consumer discretionary | null | 0.08 | null | 85.5 | United States | NASDAQ | - | null |
|-| abr | arbor realty trust reit inc | financials | null | 0.01 | null | 13.19 | United States | New York Stock Exchange Inc. | - | null |
|-| labt | labotech laboratories | health care | null | 0.5 | null | 180.79 | United States | New York Stock Exchange Inc. | - | null |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

Schema of metadata:
root
|-- ticker: string (nullable = true)
|-- company_name: string (nullable = true)
|-- sector: string (nullable = true)
|-- market_value: float (nullable = true)
|-- weight_percent: float (nullable = true)
|-- shares_quantity: integer (nullable = true)
|-- share_price: float (nullable = true)
|-- country: string (nullable = true)
|-- stock_exchange: string (nullable = true)
|-- accrual_date: string (nullable = true)
|-- record_id: double (nullable = true)

[25] def data_storing(df, output_path):
    try:
        df.coalesce(1).write.mode("overwrite").option("header", "true").csv(output_path)
```

```
AI Powered Financial Analysis.ipynb
File Edit View Insert Runtime Tools Help
Q Commands + Code + Text
[25] output_path = "processed_sec_filings"

data_storing(extracted_metadata, output_path)

Data successfully saved to processed_sec_filings
Verification: Read back 2622 rows from processed_sec_filings
Sample of stored data:
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| ticker | company_name | sector | market_value | weight_percent | shares_quantity | share_price | country | stock_exchange | accrual_date | record_id |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|-| comiab inc 1250 vesting prvt | health care | 0.02 | 0.0 | null | 0.0 | United States | NO MARKET (E.G. UNLISTED) | - | null | |
|-| comiab inc 1500 vesting prvt | health care | 0.02 | 0.0 | null | 0.0 | United States | NO MARKET (E.G. UNLISTED) | - | null |
|-| agilent technologies inc | health care | null | 0.12 | null | 149.65 | United States | New York Stock Exchange Inc. | - | null |
|-| alisa corp | materials | null | 0.02 | null | 45.47 | United States | New York Stock Exchange Inc. | - | null |
|-| american airlines group inc | industrials | null | 0.02 | null | 12.72 | United States | NASDAQ | - | null |
|-| aerom company inc | consumer discretionary | null | 0.0 | null | 13.95 | United States | New York Stock Exchange Inc. | - | null |
|-| ason inc | industrials | null | 0.01 | null | 75.32 | United States | NASDAQ | - | null |
|-| aspf | advance auto parts inc | consumer discretionary | null | 0.02 | null | 147.63 | United States | New York Stock Exchange Inc. | - | null |
|-| apple inc | information technology | null | 5.16 | null | 129.93 | United States | NASDAQ | - | null |
|-| asat | american assets trust reit inc | real estate | null | 0.0 | null | 26.5 | United States | New York Stock Exchange Inc. | Jan 01, 1970 | null |
|-| asaw | atlas air worldwide holdings inc | industrials | null | 0.01 | null | 180.8 | United States | NASDAQ | - | null |
|-| abbv | abbvie inc | health care | null | 0.76 | null | 161.61 | United States | New York Stock Exchange Inc. | - | null |
|-| abc | americansourceberg corp | health care | null | 0.07 | null | 165.71 | United States | New York Stock Exchange Inc. | - | null |
|-| abcb | ameris bancorp | financials | null | 0.01 | null | 147.14 | United States | NASDAQ | - | null |
|-| abcl | abcellera biologics inc | health care | null | 0.01 | null | 18.13 | Canada | NASDAQ | - | null |
|-| abg | ashbury automotive group inc | consumer discretionary | null | 0.01 | null | 179.25 | United States | New York Stock Exchange Inc. | - | null |
|-| abe | abe industries inc | industrials | null | 0.01 | null | 44.42 | United States | New York Stock Exchange Inc. | - | null |
|-| absh | airbnb inc class a | consumer discretionary | null | 0.08 | null | 85.5 | United States | NASDAQ | - | null |
|-| abr | arbor realty trust reit inc | financials | null | 0.01 | null | 13.19 | United States | New York Stock Exchange Inc. | - | null |
|-| labt | labotech laboratories | health care | null | 0.5 | null | 180.79 | United States | New York Stock Exchange Inc. | - | null |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

[26] !pip install -q pyspark==3.4.1 spark-nlp==5.1.3
!pip install faiss-cpu

Requirement already satisfied: faiss-cpu in /usr/local/lib/python3.11/dist-packages (1.10.0)
Requirement already satisfied: numpy<3.0,>=1.25.0 in /usr/local/lib/python3.11/dist-packages (from faiss-cpu) (2.0.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from faiss-cpu) (24.2)

[27] from pyspark.sql import SparkSession

spark_session = SparkSession.getActiveSession()

if spark_session:
    spark_session.stop()

[28] !import sparkhelp
```



```
AI_Powered_Financial_Analysis.ipynb
File Edit View Insert Runtime Tools Help

[60] print("\nRetrieved (Top 3):\n[doc[100] + '...' for doc in retrieved_docs[3]]")
print("\nLatency: {latency:.2f}s")

print("\n=== Fine-Tuning Retrieval ===")
query = "What are the risks of Amazon's $15B warehouse expansion?"
response, retrieved_docs, latency = run_rag(query, faiss_index, processed_docs, model, k=3)
print("Query: {query}")
print("Answer (k=3):\n{response}")
print(f"Retrieved (Top 3):\n[doc[100] + '...' for doc in retrieved_docs[3]]")
print("\nLatency: {latency:.2f}s")

=== Testing Diverse Financial Queries ===
Device set to use cpu

Query: What are the risks of Amazon's $15B warehouse expansion?
Answer:
I think there is still a huge downside risk for the US stock market, despite the exemption for electronics Amazon Cancels Inventory Orders From China After Tariffs ECB cuts rates again to help economy weather erratic U.S. trade policy
Retrieved (Top 3):
['Amazon considers $15 billion warehouse expansion plan, Bloomberg News reports...', 'I think there is still a huge downside risk for the US stock market, despite the exemption for elect...', 'Amazon Cancels Inventory Orders From China After Tariffs...']
Latency: 6.86s
Device set to use cpu

Query: What is the impact of Trump's tariffs on inflation?
Answer:
controlling inflation and supporting economic growth.
Retrieved (Top 3):
['HAWKSWATCH: BIG TARIFFS, BIGGER DEFICITS...', 'Powell indicates tariffs could pose a challenge for the Fed between controlling inflation and support...', 'What is this gonna look like after 100% tariffs? ...']
Latency: 3.46s
Device set to use cpu

Query: Apple is acquiring a startup in New York.
Answer:
No.
Retrieved (Top 3):
['Apple redirects iPhone production to India amidst high China tariffs...', 'Bad Apple...', 'Apple is charting flights for 600 tons of iPhones from India...']
Latency: 1.66s

=== Fine-Tuning Retrieval ===
Device set to use cpu
Query: What are the risks of Amazon's $15B warehouse expansion?
Answer (k=3):
downside risk for the US stock market
Retrieved (Top 3):
['Amazon considers $15 billion warehouse expansion plan, Bloomberg News reports...', 'I think there is still a huge downside risk for the US stock market, despite the exemption for elect...', 'Amazon Cancels Inventory Orders From China After Tariffs...']
Latency: 1.69s

[61] print("\n=== Optimize and Finalize ===")
nlist = 100
quantizer = faiss.IndexLatIP(dimension)
optimized_index = faiss.IndexIVFlat(quantizer, dimension, nlist, faiss.METRIC_INNER_PRODUCT)
optimized_index.train(embeddings)
optimized_index.add(embeddings)

response, retrieved_docs, latency = run_rag(query, optimized_index, processed_docs, model)
print("Query (Optimized): {query}")
print("Answer:\n{response}")
print("\nLatency: {latency:.2f}s")

=== Optimize and Finalize ===
Device set to use cpu
Query (Optimized): What are the risks of Amazon's $15B warehouse expansion?
Answer:
ECB cuts rates again to help economy weather erratic U.S. trade policy Companies with larger fixed USD debt?
Latency: 3.55s
```

```
AI_Powered_Financial_Analysis.ipynb
File Edit View Insert Runtime Tools Help

downside risk for the US stock market
Retrieved (Top 3):
['Amazon considers $15 billion warehouse expansion plan, Bloomberg News reports...', 'I think there is still a huge downside risk for the US stock market, despite the exemption for elect...', 'Amazon Cancels Inventory Orders From China After Tariffs...']
Latency: 1.69s

[63] print("\n=== Optimize and Finalize ===")
nlist = 100
quantizer = faiss.IndexLatIP(dimension)
optimized_index = faiss.IndexIVFlat(quantizer, dimension, nlist, faiss.METRIC_INNER_PRODUCT)
optimized_index.train(embeddings)
optimized_index.add(embeddings)

response, retrieved_docs, latency = run_rag(query, optimized_index, processed_docs, model)
print("Query (Optimized): {query}")
print("Answer:\n{response}")
print("\nLatency: {latency:.2f}s")

=== Optimize and Finalize ===
Device set to use cpu
Query (Optimized): What are the risks of Amazon's $15B warehouse expansion?
Answer:
ECB cuts rates again to help economy weather erratic U.S. trade policy Companies with larger fixed USD debt?
Latency: 3.55s

Colab paid products - Cancel contracts here
```