

Team name: Finance Bull 🐮

Chu-Huan Huang, 1934-2084-30

1. Team member names and USC IDs:

- Chu-Huan Huang, 1934208430
- Chung-Sheng Chang, 1676324151
- Jiaying Wang, 3018182967

2. The domains we chose:

- Stock trading
- Finance
- Economics

3. Shortlisted data options, reasons, and the data sources:

- Stock Price-Related Data:
 - Types:
 - Stock market Index, stock prices, news sentiment, etc.
 - Reasons:
 - We can analyze the stock market trends using the real-time and historical data mentioned above, along with insights from investors on stock market forums like Reddit. By examining various tools and models, we can offer valuable information to users, such as predictions of future trends and suggestions for investment directions.
 - Data Sources:
 - Yahoo Finance (<https://finance.yahoo.com/>):

Yahoo Finance offers comprehensive financial information, including real-time stock quotes, news, historical data, and company financials. We can utilize the API and the downloaded CSV files of a company's stock prices or news text.

Sample excerpt:

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
2024-01-2	194.2683	195.3132	192.1686	193.2234	54822100	0	0
2024-01-2	193.3229	193.8105	191.0043	191.4819	44594000	0	0
2024-01-2	191.0739	191.263	188.6558	190.7953	47145600	0	0
2024-01-3	190.0092	190.865	186.5561	187.1233	55859400	0	0
2024-01-3	186.1282	186.1879	183.4513	183.501	55467800	0	0
2024-02-0	183.093	186.0386	182.9239	185.949	64885400	0	0
2024-02-0	178.9832	186.4167	178.3761	184.944	1.03E+08	0	0
2024-02-0	187.2328	188.3274	184.934	186.765	69668800	0	0
2024-02-0	185.949	188.3871	185.8595	188.3771	43490800	0	0

- r/stock (<https://www.reddit.com/r/stocks/>):

r/stocks is a Reddit community where users discuss stocks, market news, and investing strategies, as well as share educational resources. We can analyze text retrieved from here to gauge investors' sentiment.

Sample excerpt:

Bank of Japan Raised its short-term policy rate from 0.25% to 0.5%

Broad market news

Per [Reuters](#): BOJ Governor Kazuo Ueda said the central bank will keep raising interest rates as wage and price increases broaden, adding that there was scope to push up borrowing costs further before they reach levels deemed neutral to the economy.

But he offered few clues on the timing and pace of future rate hikes, saying the decision will be based on how soon Japan will see trend inflation sustainably hit the BOJ's target.

"We don't have any preset idea. We'll make a decision at each policy meeting by looking at economic and price developments as well as risks," he told a press conference after the policy decision.

- Finance-Related Data
 - Types:
 - Corporate financial statements, earnings reports, bond yields, credit ratings, and other related documents.
 - Reasons:
 - By examining corporate financial statements and earnings reports, we can analyze a company and provide insights to the users of our chatbot.
 - Data Sources:
 - SEC EDGAR (<https://www.sec.gov/>):

SEC EDGAR is a reliable source for accessing corporate financial reports, including quarterly and annual filings, 10-Ks, and 10-Qs.

Sample excerpt:

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 10-K

(Mark One)

☒ ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934
For the fiscal year ended September 28, 2024

☐ TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934
For the transition period from _____ to _____
Commission File Number: 001-38743


Apple Inc.
(Exact name of Registrant as specified in its charter)

California
(State or other jurisdiction of incorporation or organization)

One Apple Park Way
Cupertino, California
(Address of principal executive offices)

942404110
(I.R.S. Employer Identification No.)

95014
(Zip Code)

(408) 996-1010

- Discounting Cash Flow (<https://discountingcashflows.com/>):

DCF provides organized financial data and the complete history of financial statements, including links to a company's original SEC filings (10-K and 10-Q), which can be downloaded as Excel files.

Sample excerpt:

Period End	2024	09-2 2023	09-3 2022	09-2 2021	09-2 2020	09-2 2019
Report Filed	2024-11-01	2023-11-01	2022-10-02	2021-10-02	2020-10-02	2019-10-01
Total Current Assets	152,987	143,566	135,405	134,836	143,713	143,713
Cash and Cash Equivalents	65,171	61,555	48,304	62,639	90,943	90,943
Cash & Equivalents	29,943	29,965	23,646	34,940	38,016	38,016
Short-Term Investments	35,228	31,590	24,658	27,699	52,927	52,927
Receivables	66,243	60,985	60,932	51,506	37,445	37,445
Inventory	7,286	6,331	4,946	6,580	4,061	4,061
Other Current Assets	14,287	14,695	21,223	14,111	11,264	11,264
Total Assets	364,980	352,583	352,755	351,002	323,888	323,888
Total Non-Current Assets	211,993	209,017	217,350	216,166	180,175	180,175
Property, Plant & Equipment	45,680	54,376	52,534	49,527	45,336	45,336
Goodwill	0.00	0.00	0.00	0.00	0.00	0.00
Goodwill	0.00	0.00	0.00	0.00	0.00	0.00
Intangible Assets	0.00	0.00	0.00	0.00	0.00	0.00
Long-Term Investments	91,479	100,544	120,805	127,877	100,887	100,887
Deferred Tax Assets	19,499	17,852	15,375	0.00	0.00	0.00
Other Long-Term Assets	55,335	36,245	28,636	38,762	33,952	33,952
Total Current Liabilities	176,392	145,308	153,982	125,481	105,392	105,392
Accounts Payable	68,960	62,611	64,115	54,763	42,296	42,296
Notes Payable	22,511	17,382	22,773	17,141	15,229	15,229
Tax Payable	26,601	8,819	6,552	0.00	0.00	0.00
Deferred Tax Liabilities	8,249	8,061	7,912	7,612	6,643	6,643
Other Current Liabilities	50,071	48,435	52,630	45,965	41,224	41,224

- Economics-Related Data

- Types:

- GDP, inflation rates, unemployment rates, global trade statistics, energy prices, etc.
- Reasons:
 - Macroeconomic indicators, such as GDP, inflation, and unemployment, are essential for assessing the overall health of an economy. Trade statistics provide insights into international market trends, while energy price data helps us understand the economic impact of fluctuations in global energy markets. By analyzing these datasets, we can offer a comprehensive picture of economic conditions and trends to users of the chatbot.
- Data Sources:
 - World Bank Open Data (<https://data.worldbank.org/>):

The World Bank provides global economic data, including GDP, trade statistics, and demographic information. These CSV-type datasets can be used for country-level economic analysis and forecasting.

sample excerpt:

Country N	Country C	Indicator N	Indicator C	1960	1961	1962	1963	1964	1965	1966	1967	1968
Aruba	ABW	GDP (curr	NY.GDP.MKTP.CD									
Africa East	AFE	GDP (curr	NY.GDP.M	2.42E+10	2.5E+10	2.71E+10	3.18E+10	3.03E+10	3.38E+10	3.69E+10	3.85E+10	4.18E+10
Afghanistan	AFG	GDP (curr	NY.GDP.MKTP.CD									
Africa Wes	AFW	GDP (curr	NY.GDP.M	1.19E+10	1.27E+10	1.36E+10	1.45E+10	1.58E+10	1.69E+10	1.8E+10	1.65E+10	1.7E+10
Angola	AGO	GDP (curr	NY.GDP.MKTP.CD									
Albania	ALB	GDP (curr	NY.GDP.MKTP.CD									
Andorra	AND	GDP (curr	NY.GDP.MKTP.CD									
Arab World	ARB	GDP (curr	NY.GDP.MKTP.CD	2E+10	2.03E+10	2.24E+10	2.48E+10	2.68E+10	2.83E+10	3.06E+10	3.5E+10	
United Ara	ARE	GDP (curr	NY.GDP.MKTP.CD									
Argentina	ARG	GDP (curr	NY.GDP.MKTP.CD		2.45E+10	1.83E+10	2.56E+10	2.83E+10	2.86E+10	2.43E+10	2.64E+10	

4. Script Description:

This Python script automates collecting and organizing financial data related to Apple Inc. from various sources. It accomplishes three main tasks:

First, the script retrieves historical stock price data for the past year using the yfinance library and saves it as a CSV file. It also displays basic information about this dataset, including the first ten rows, its dimensions, and any missing values.

Second, the script scrapes news articles related to AAPL from Yahoo Finance News. Utilizing web scraping techniques with the requests and BeautifulSoup libraries, it extracts key information such as headlines, sources, and links. This data is saved to a CSV file; the script similarly displays important details about this dataset.

Finally, the script processes a PDF document, likely an annual report, using the `fitz` from `PyMuPDF` library. It extracts the text content from each page and saves it as a text file. Before saving, the script prints the first 100 characters and the total character count of the extracted text, providing a quick overview of the content.

5. **What Chatbot missing:**

Due to several fundamental limitations, existing chatbots often encounter significant challenges in replicating human conversation. A primary issue is their inability to grasp context and understanding fully. Typically, chatbots depend on keyword matching and pre-programmed responses, which restricts their capability to handle nuanced language, sarcasm, or ambiguous inquiries. They frequently struggle to maintain context throughout extended interactions, resulting in responses that can be irrelevant or repetitive. Additionally, a restricted knowledge base can impede a chatbot's ability to provide thorough or accurate information, especially regarding specialized or evolving topics.

My datasets have the potential to significantly enhance chatbot performance in various ways. The structured stock data establishes a factual foundation for addressing queries about historical performance, trends, and key financial metrics. The news articles offer real-time context and sentiment analysis capabilities, enabling the chatbot to discuss recent events and their potential impacts on companies. Lastly, the unstructured text from financial reports provides detailed insights into a company's financials, strategy, and risk factors, allowing the chatbot to tackle more complex and nuanced questions. By integrating these diverse data sources, the chatbot can deliver more accurate, comprehensive, and contextually relevant responses, resulting in more engaging and human-like interactions.

6. **Name & USC-ID:** Chu-Huan Huang, 1934-2084-30