

Quarterly Earnings Predictor

Based on News Articles

By Matt Elmajian

Problem

- Wall Street Analysts provide Estimated Earnings for Publicly traded companies
- Companies announce earnings approximately 1 month after closing their books
- Can articles posted within one week of a company's earnings announcement predict whether or not they will outperform or underperform their estimate

Data Wrangling

- AlphaVantage API
 - Provides API to pull companies' earnings estimates and actuals dating back to 1990
 - Provides API to pull news articles dating back to 2022
- Pull data on 30 companies over their last 10 earnings reports and associated articles
 - Companies: ['AAPL', 'NVDA', 'MSFT', 'AVGO', 'META', 'ORCL', 'CRM', 'AMD', 'UBER', 'ADBE', 'IBM', 'QCOM', 'MU', 'ANET', 'GOOG', 'TSM', 'TSLA', 'NFLX', 'BABA', 'CSCO', 'SONY', 'SHOP', 'PLTR', 'INTU', 'ADP', 'DELL', 'ABNB', 'SPOT', 'PYPL', 'CRWD']
- Use BeautifulSoup to webscrape actual articles based on url's provided

API Outputs

- Earnings API Output

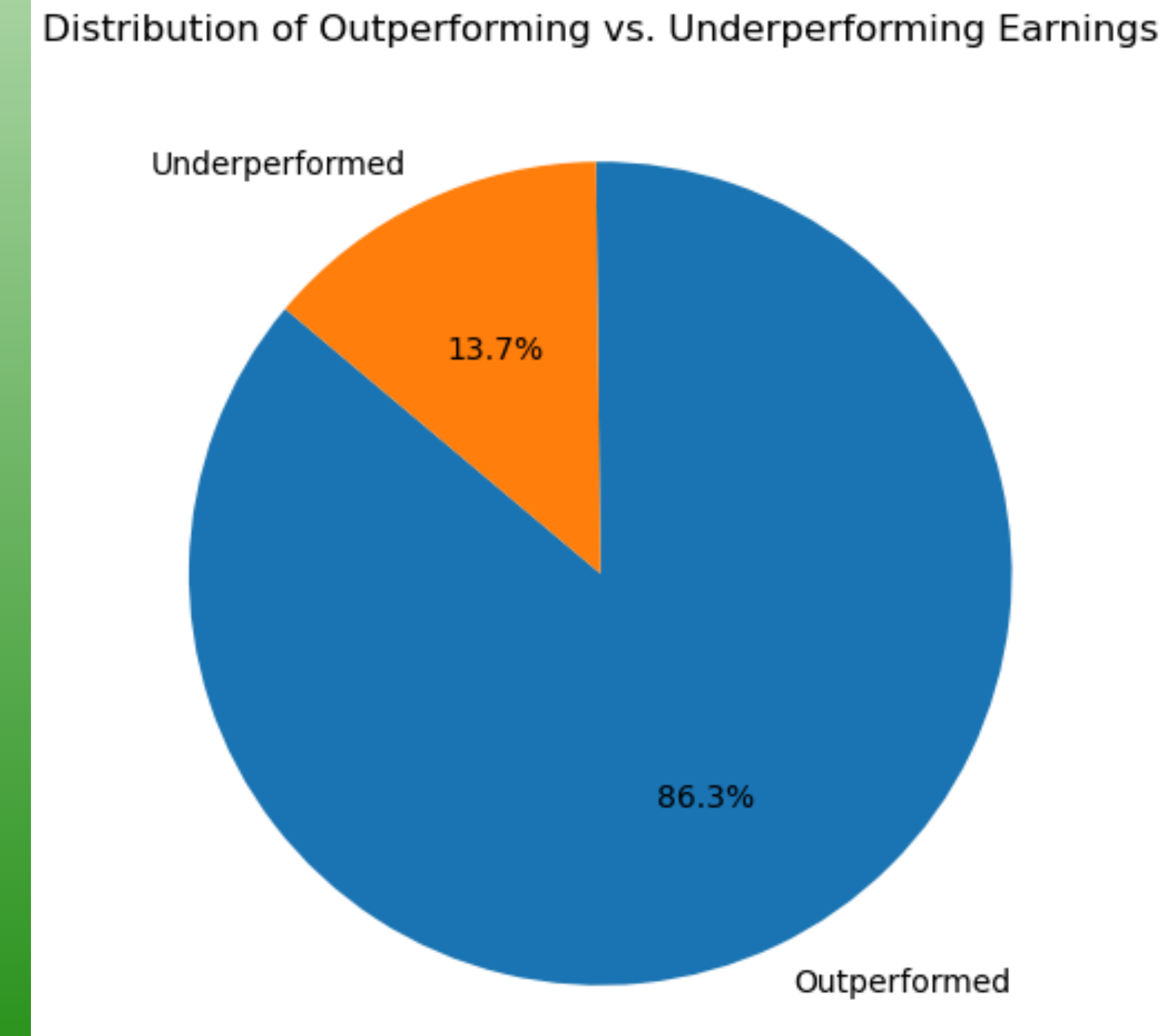
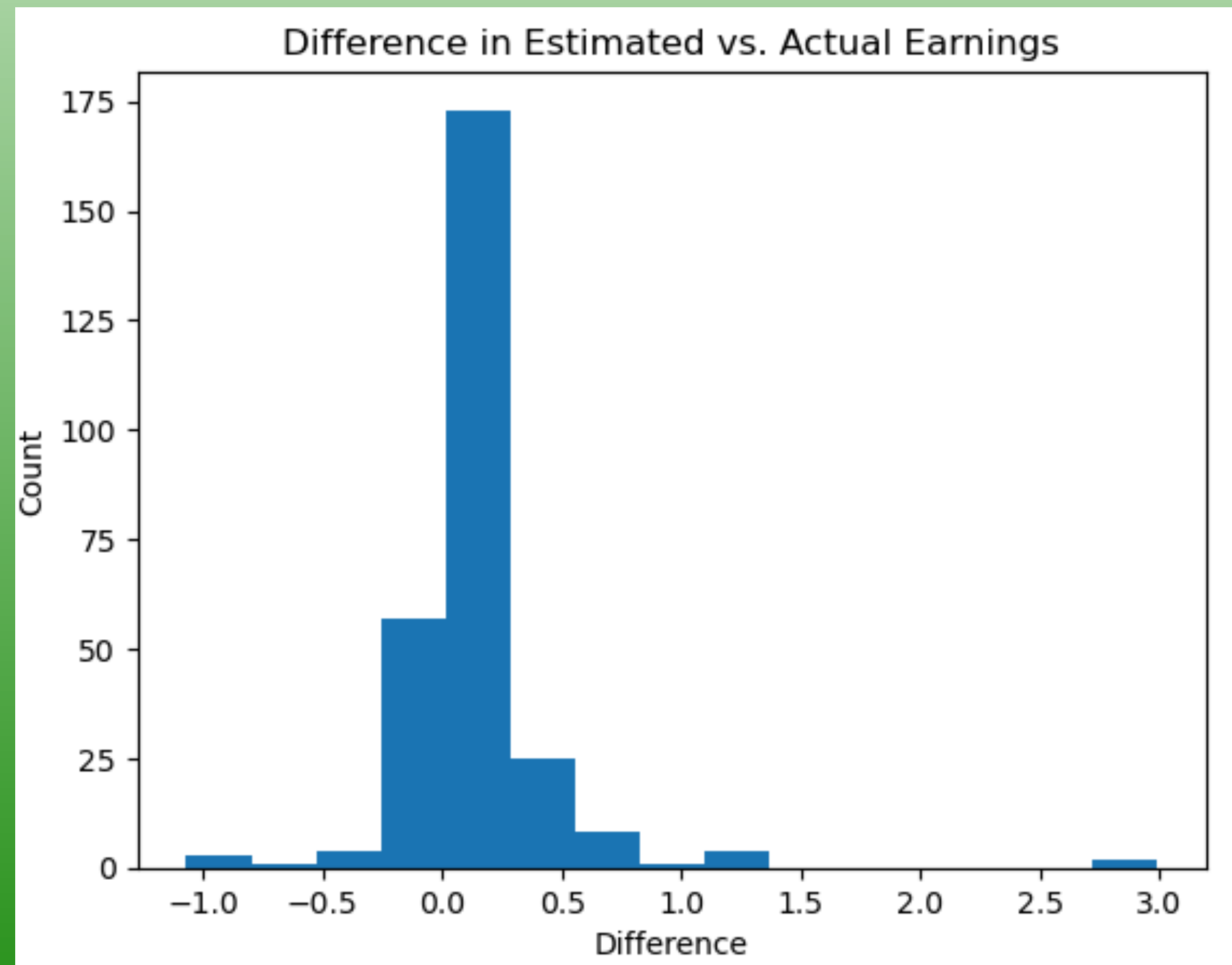
	fiscalDateEnding	reportedDate	reportedEPS	estimatedEPS	surprise	surprisePercentage	reportTime	ticker
0	2024-06-30	2024-08-01	1.4	1.35	0.05	3.7037	post-market	AAPL
1	2024-03-31	2024-05-02	1.53	1.5	0.03	2	post-market	AAPL
2	2023-12-31	2024-02-01	2.18	2.1	0.08	3.8095	post-market	AAPL
3	2023-09-30	2023-11-02	1.46	1.39	0.07	5.036	post-market	AAPL
4	2023-06-30	2023-08-03	1.26	1.19	0.07	5.8824	post-market	AAPL

- News API Output

	title	url	time_published	authors	summary	banner_image	source
0	Advanced Energy Q2 Earnings Top Estimates, Rev...	https://www.benzinga.com/news/earnings/24/07/4...	20240731T201919	[Zacks]	Advanced Energy Industries AEIS reported non-G...	https://staticx-tuner.zacks.com/images/charts/...	Benzinga
1	AMD Q2 Earnings Beat Estimates, Shares Up on S...	https://www.zacks.com/stock/news/2313423/amd-q...	20240731T181200	[Zacks Investment Research]	AMD's second-quarter 2024 results reflect robu...	https://staticx-tuner.zacks.com/images/article...	Zacks Commentary

EDA - Target Variable

- Over the last two years, many of the companies tested outperformed their estimates.
- 86.3% at or above estimates



Data Wrangling

- Cleanings & Preprocessing:
 - Remove numbers
 - Remove punctuation
 - Remove Emojis
 - Lemmatize words
- Used TFIDF & Count Vectorizer

