



DEV10 CAPSTONE

# **THE FINANCE INDUSTRY AND THE STOCK MARKET**

DAN ANGELINA  
NICOLE BOUTDARA  
MICHAEL DEEHAN  
CAITLIN RYAN

# Introduction

- The stock market experiences changes throughout the economic cycle and can be affected by major events.
- The finance industry is one of the largest industries trading in the stock market.
- In our research, we observe historical stock prices and utilize machine learning to understand and explain the market and its changes.

# Agenda

**01** Exploratory Questions

**02** Data Sources

**03** Industry Overview

**04** Data Findings

**05** Machine Learning Model

# EXPLORATORY QUESTIONS

01

Which sub-category of the finance industry leads in sales/revenue?

02

How do companies perform compared to each other in the stock market?

03

How do major events impact stock performance? How do expectations affect these outcomes?

04

How does the industry tend to perform in comparison to the S&P 500 in terms of stock price changes?

05

How accurately can we predict whether a stock will perform better or worse relative to the S&P 500?

# Data Sources

- Summary statistics of the finance industry from the United States Census Bureau
- Historical stock price and other financial data from *The Wall Street Journal* and Yahoo Finance
  - Over 14 years of daily stock price data covering the Great Recession, 4 presidential elections, and the current state of the economy during the Coronavirus pandemic
  - Stock price data includes the daily opening, closing, high, and low prices, as well as the volume of shares traded

# Data Sources

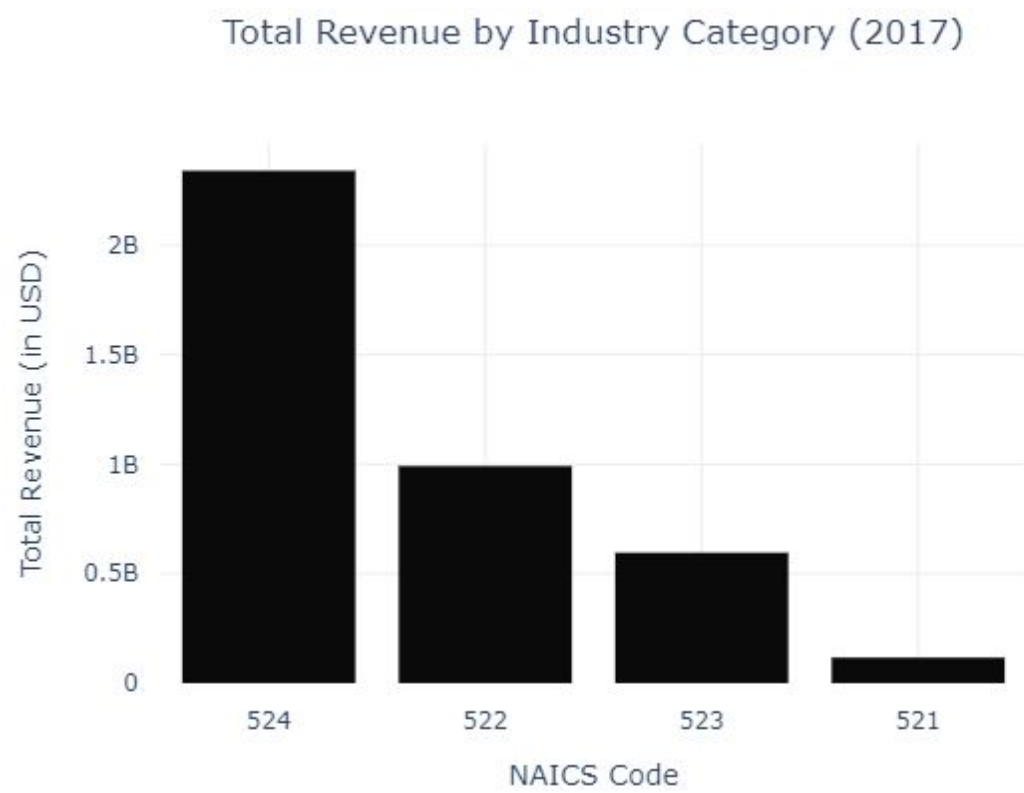
- Six stocks in the financial/banking industry traded on the New York Stock Exchange
- Differing levels of market capitalization to get a diverse look at the industry

Company Name	Ticker Symbol	Market Capitalization	Size Classification
JP Morgan Chase & Co.	JPM	\$495.83B	Large
Goldman Sachs Group Inc.	GS	\$135.71B	Large
Discover Financial Services	DFS	\$34.23B	Medium
Synovus Financial Corp.	SNV	\$7.17B	Medium
New York Community Bancorp Inc.	NYCB	\$5.86B	Small
Bank of Hawaii Corp.	BOH	\$3.52B	Small

# Extract, Transform, Load

- Extraction
  - We downloaded all of our datasets as CSV files and stored them in our team's *group2\_capstone* container.
- Transformation
  - Luckily, our datasets were quite clean and void of any null or unpleasant values.
  - For our Census data, we utilized Power Query to remove columns, change headers, remove duplicate and unnecessary rows, and create new columns.
  - For our stock and S&P 500 data, we primarily worked with Pandas to transform our data. Using Pandas, we converted column types, merged our S&P 500 and stock data frames, and added new columns for percent changes, exponentially weighted moving average, and market cap.
- Load
  - After sending our data frame through our Producer (one row at a time) to our Consumer, we saved our data
  - After compiling our data, we needed to send it to SQL. For this step, we needed to create an ID column for each table, set our primary keys, and once again adjust our column types,

# Industry Overview



**54%** Insurance Carriers  
\$2,340,445,842

**29%** Credit Intermediation  
\$1,286,777,214

**14%** Investment Activities  
\$595,895,851

**3%** Monetary Authorities  
\$116,892,000

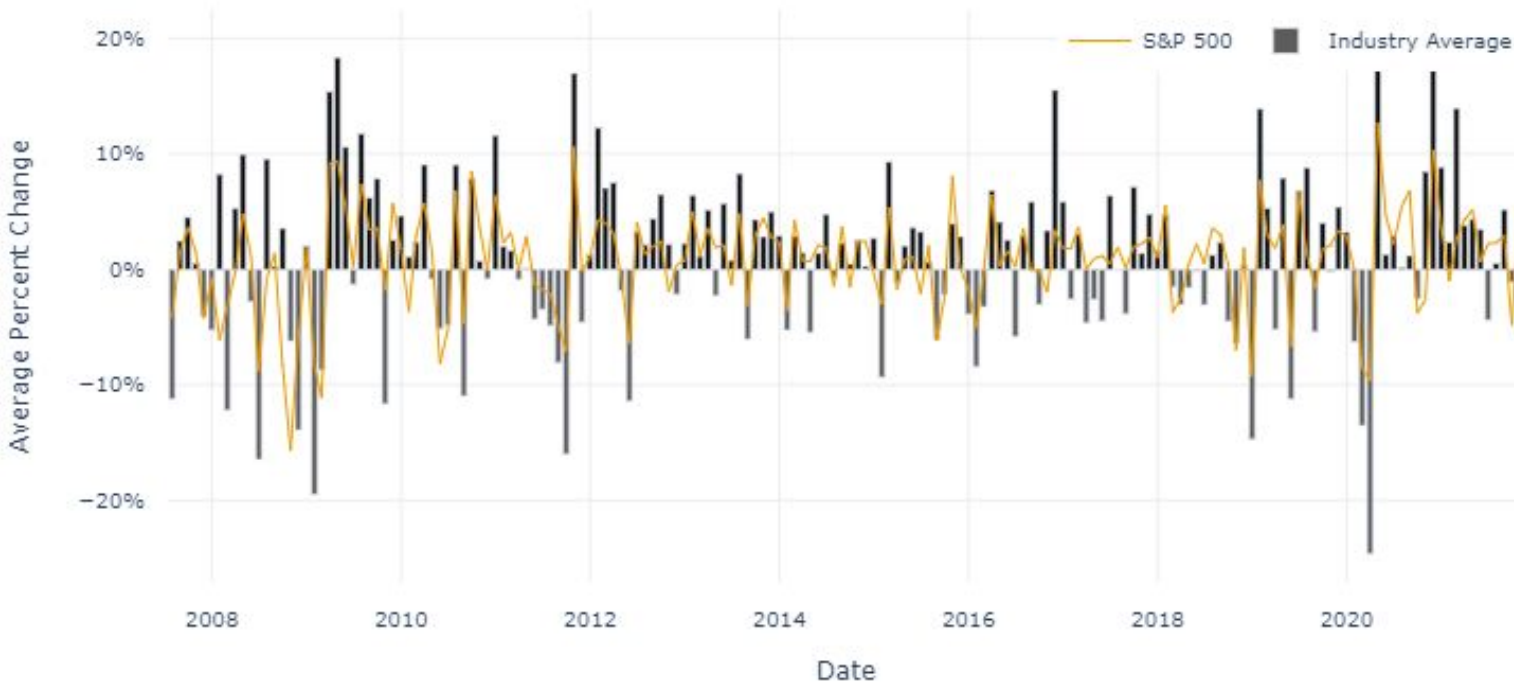
# **DATA FINDINGS**



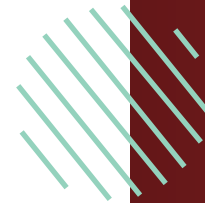


- By October 2021, positions generally follow the market capitalization breakdown.
- While Goldman Sachs has the highest closing price value, JP Morgan has a much larger market capitalization due to having far more shares publicly traded.
- Prices fell during the Great Recession and the Coronavirus pandemic.

Monthly Percent Change in Closing Prices Compared to the S&P 500 (2007–2021)



- The industry generally has higher highs and lower lows than the S&P 500.
- No noticeable patterns with the presidential elections.
- Consecutive negative values in 2011 presented an opportunity for further research.



# Machine Learning Model

How accurately can we predict whether a stock price will perform better or worse than the S&P 500?

- What data to include?
  - Problems: need an output column for ML, highly correlated columns, not all columns are numeric
  - Solutions: create new columns, create dummy variables
- New columns
  - Exponentially weighted moving average of closing price column grouped by company
  - Comparison column between a stock's percent change and the S&P's percent change is binary for output column
  - Use previous days differences between a stock's percent change and the S&P's percent change
  - Categories based on market capitalization to be turned into dummy variable

# Model Results

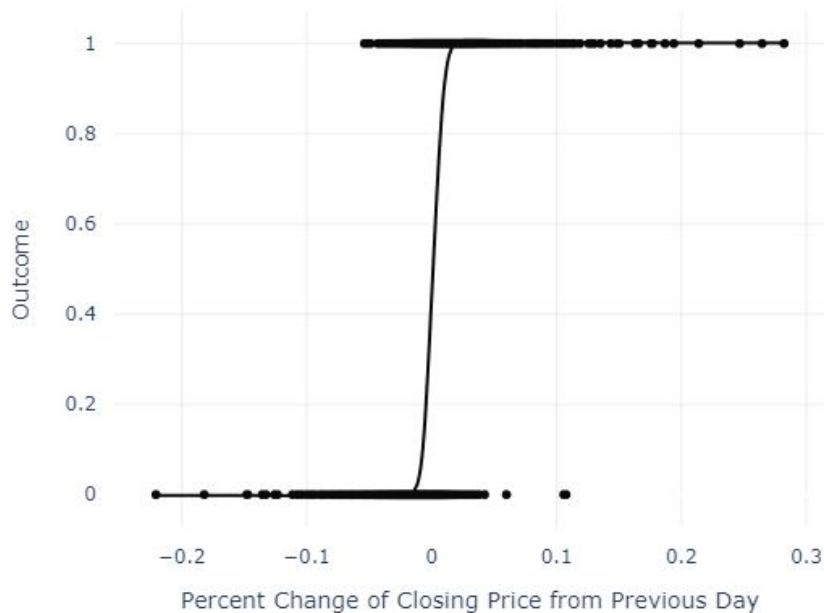
**0.819**

Model Score

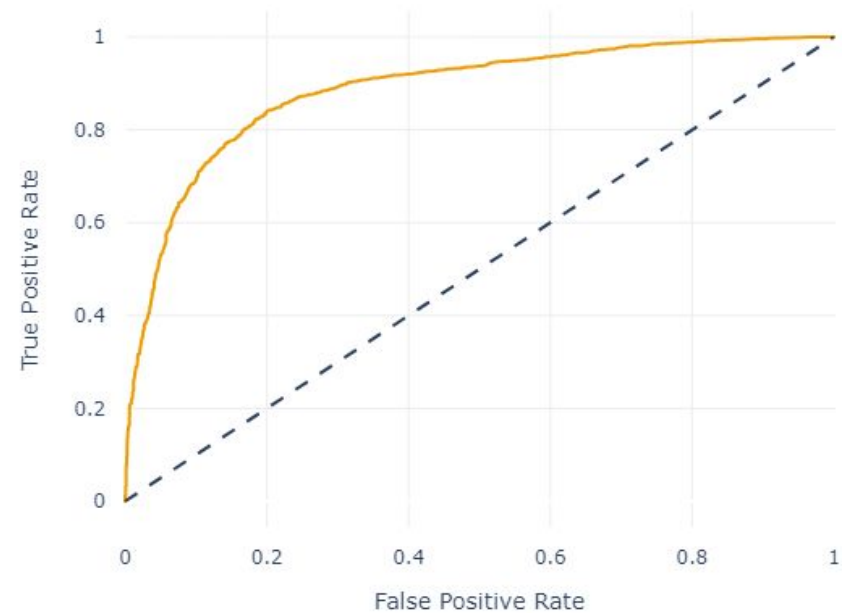
**0.638**

Matthews Correlation  
Coefficient

Logistic Regression Plot of Machine Learning Testing Data



Receiver Operating Characteristic



Anything to be done to improve the model?

# Conclusion

- Insurance carriers generate the largest share of revenue within the industry.
- Major events typically follow expected results in the market except in the case of presidential elections.
- The finance industry follows similar trends to the S&P 500, though with more extreme highs and lows.
- The finance industry is performing well financially in its current state, especially in the context of the pandemic.



**THANK  
YOU**