



Agenda



01

Problem

02

Data Collection

03

EDA

04

ML Model

05

What worked
vs. could be
Improved

06

Q&A



Problem

Goal:

Provide precise analysis for Fortune 500 companies
Predicting profit using financial data (stock price, revenue, industry, sector etc.)

Importance:

Help investor make informed decisions
Allow companies to allocate resources efficiently to optimize the profit



Data Collection

Initial Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Rank	Name	Ticker	Sector	Industry	Revenue	Profit	m1-o	m1-c	m2-o	m2-c	m3-o	m3-c	m4-o	m4-c	m5-o	m5-c
2	1	Walmart	WMT	Consumer Defensive	Discount Stores	523,964.00	14,881.00	39.62	38.1633	38.3	35.8933	35.8667	37.8733	37.3833	40.5167	40.4933	41.3533
3	2	Amazon.com	AMZN	Consumer Cyclical	Internet Retail	280,522.00	11,588.00	93.75	100.436	100.53	94.1875	95.3245	97.486	96.6485	123.7	116.84	122.1185
4	3	Exxon Mobil	XOM	Energy	Oil & Gas Integrated	264,938.00	14,340.00	70.24	62.12	61.38	51.44	52.59	37.97	36.86	46.47	45.63	45.47
5	4	Apple	AAPL	Technology	Consumer Electronic	260,174.00	55,256.00	74.06	77.3775	76.075	68.34	70.57	63.5725	61.625	73.45	71.5625	79.485
6	5	CVS Health	CVS	Healthcare	Healthcare Plans	256,776.00	6,634.00	74.65	67.82	68.4	59.18	60.67	59.33	57.75	61.55	60.79	65.57
7	6	Berkshire Hathaway	BRKA			254,616.00	81,417.00										
8	7	UnitedHealth Group	UNH	Healthcare	Healthcare Plans	242,155.00	13,839.00	293.98	272.45	275.13	254.96	257.34	249.38	238.69	292.47	288.39	304.85
9	8	McKesson	MCK	Healthcare	Medical Distribution	214,319.00	34	139.11	142.61	143.03	139.86	141.57	135.26	130.46	141.25	139.13	158.67
10	9	AT&T	T	Communication Ser	Telecom Services	181,193.00	13,903.00	51.8478	49.8089	50.259	46.6313	46.8299	38.5946	37.5751	40.3423	40.7395	40.8586
11	10	AmerisourceBergen	ABC			179,589.10	855.4										

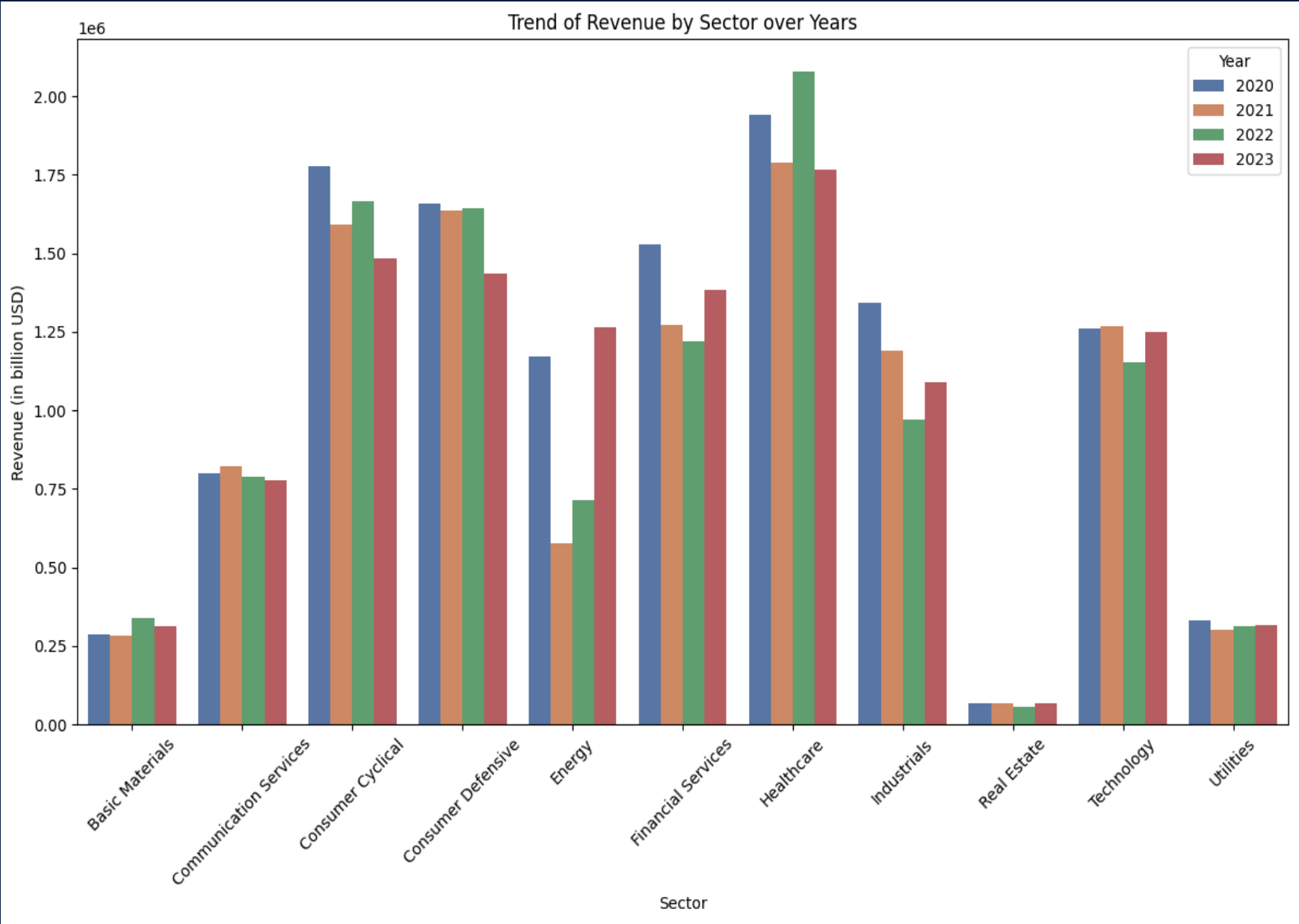
Cleaned Data

	Rank	Name	Ticker	Sector	Industry	Revenue	Profit	m1-o	m1-c	m2-o	...	m8-o	m8-c	m9-o	m9-c	m10-o	m10-c	m11-o	m11-c	m12-o	m12-c
0	1	Walmart	WMT	Consumer Defensive	Discount Stores	523964.0	14881.0	39.62	38.1633	38.300	...	43.3033	46.2833	47.0667	46.6367	46.9333	46.2500	46.8633	50.930	51.200	48.0500
1	2	Amazon.com	AMZN	Consumer Cyclical	Internet Retail	280522.0	11588.0	93.75	100.4360	100.530	...	159.0255	172.5480	174.4790	157.4365	160.4000	151.8075	153.0870	158.402	159.425	162.8465
2	3	Exxon Mobil	XOM	Energy	Oil & Gas Integrated	264938.0	14340.0	70.24	62.1200	61.380	...	42.0500	39.9400	39.7500	34.3300	33.7900	32.6200	33.1400	38.130	38.960	41.2200
3	4	Apple	AAPL	Technology	Consumer Electronics	260174.0	55256.0	74.06	77.3775	76.075	...	108.2000	129.0400	132.7600	115.8100	117.6400	108.8600	109.1100	119.050	121.010	132.6900
4	5	CVS Health	CVS	Healthcare	Healthcare Plans	256776.0	6634.0	74.65	67.8200	68.400	...	63.4500	62.1200	62.0100	58.4000	58.4300	56.0900	56.8700	67.790	68.390	68.3000

- Data Collection:
 - GitHub repository
 - Yfinance API (Industry, Sector)
 - Web Scrapping
 - Polygon API (Stocks)
 - Government data (.csv files)

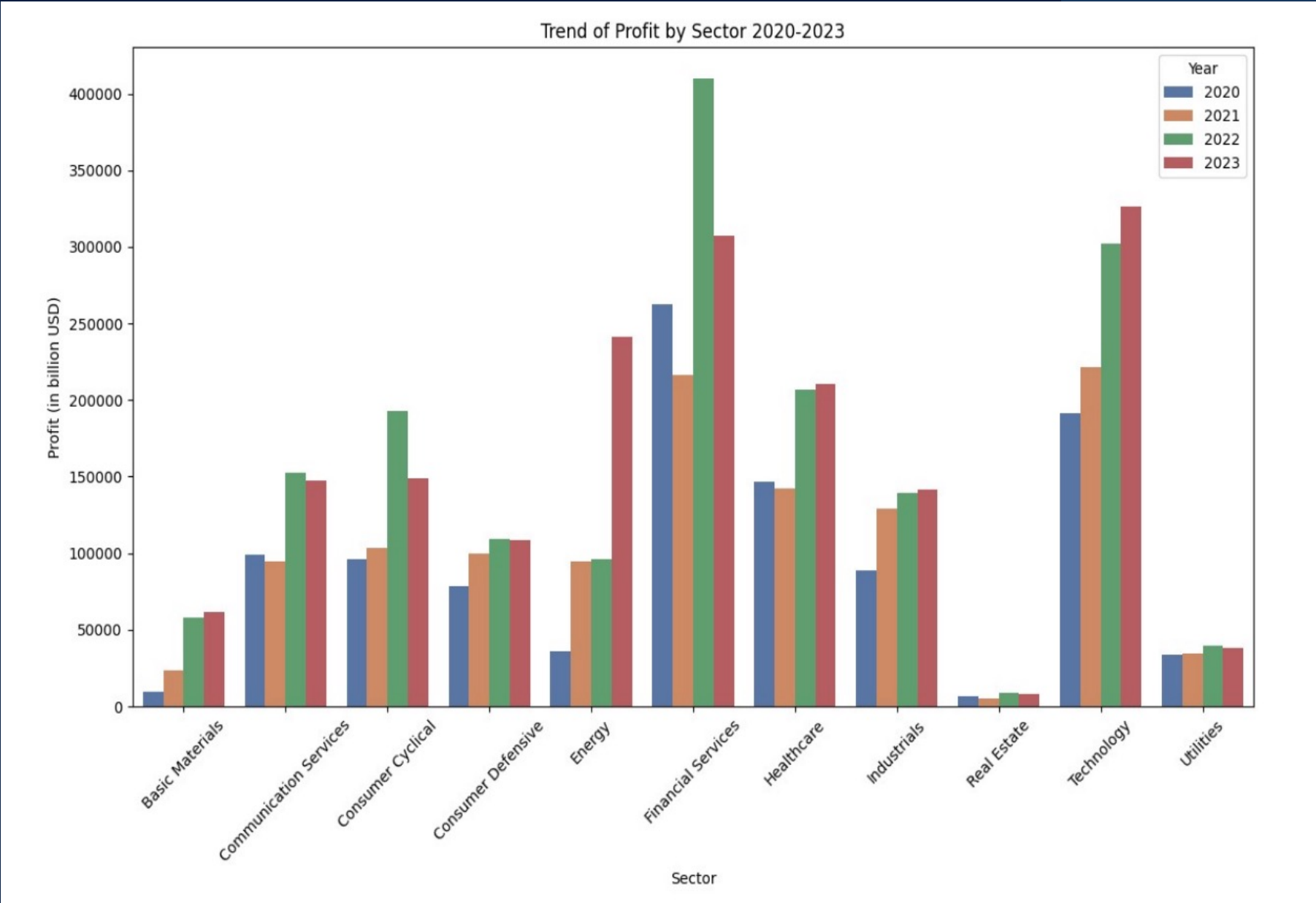
- Data Type:
 - Numerical - Continuous & Discrete
 - Categorical - Nominal
 - Text/Strings

Revenue by sector



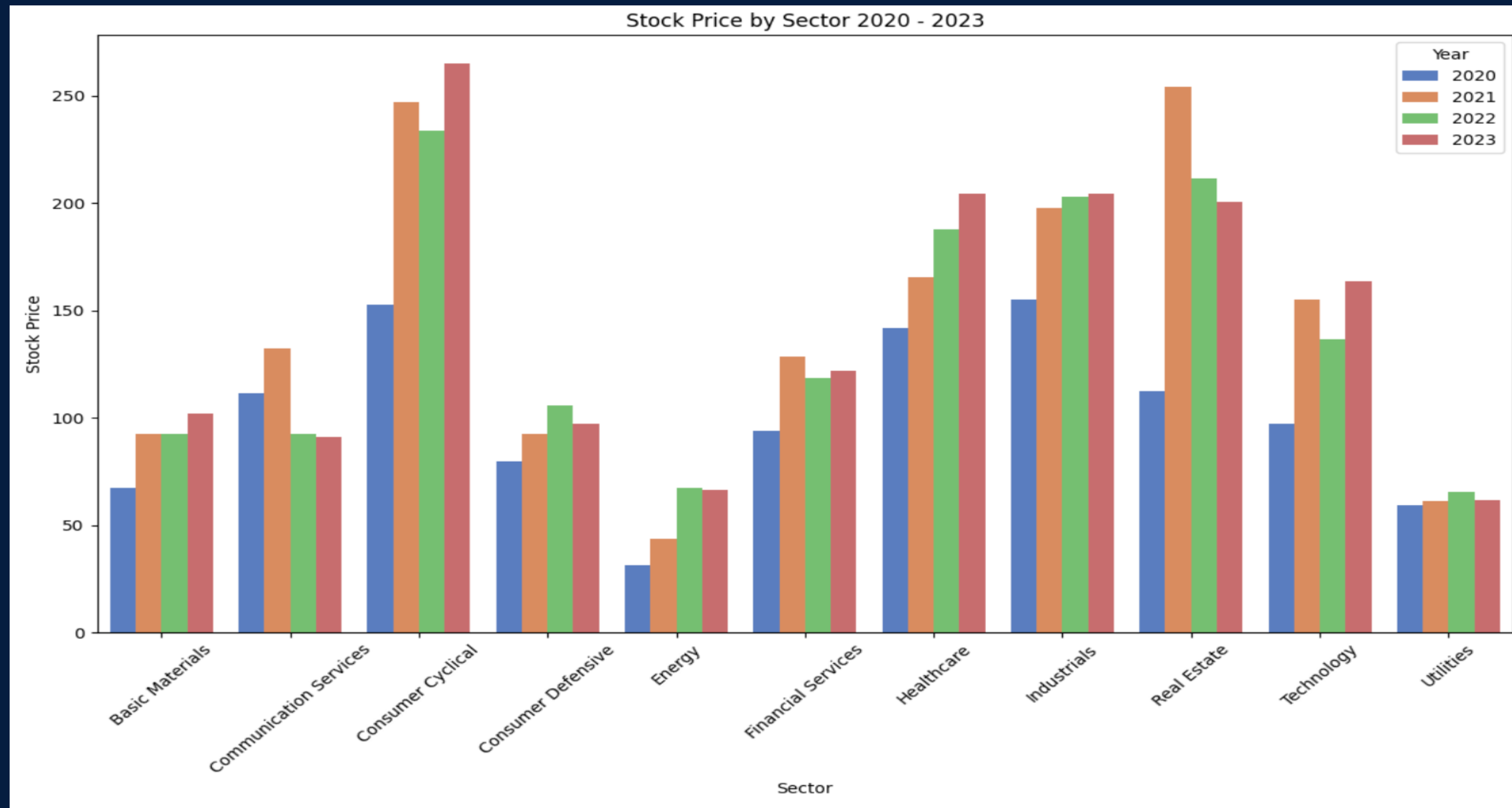
The trends have revealed sustained increase in Technology and Healthcare, demonstrates strong economic dynamics and the possibility for continuous expansion.

Profit by sector



This analysis is crucial for informing long-term strategic investments and policy-making in the IT sector, as it explores the sector's sustainability and its larger economic impact.

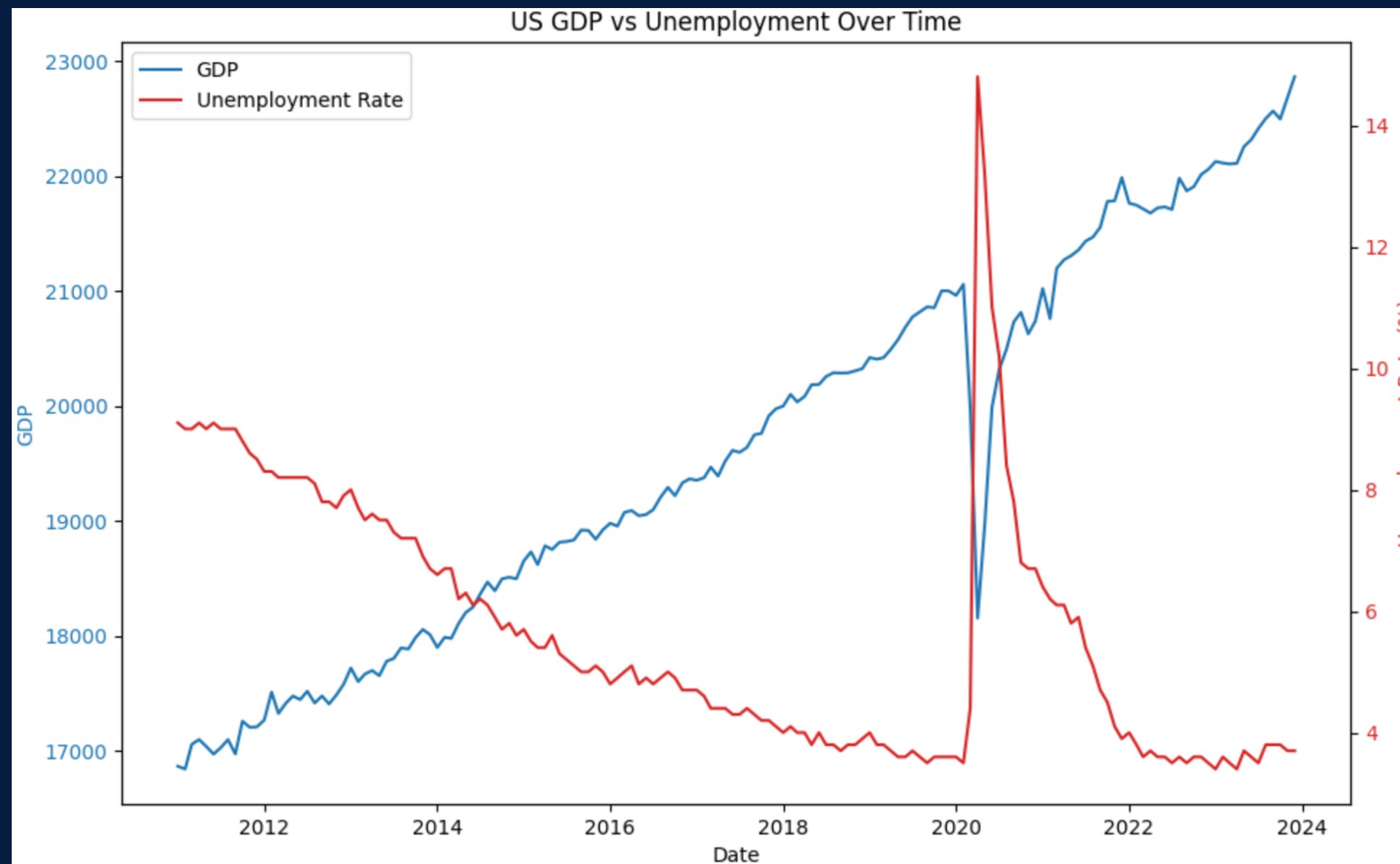
Stock Price by sector



Visualizes the mean stock prices across different sectors from 2020 - 2023

Employed Random Forest Regression to predict stock prices for the year 2024, training the model on data spanning from 2020 to 2023.

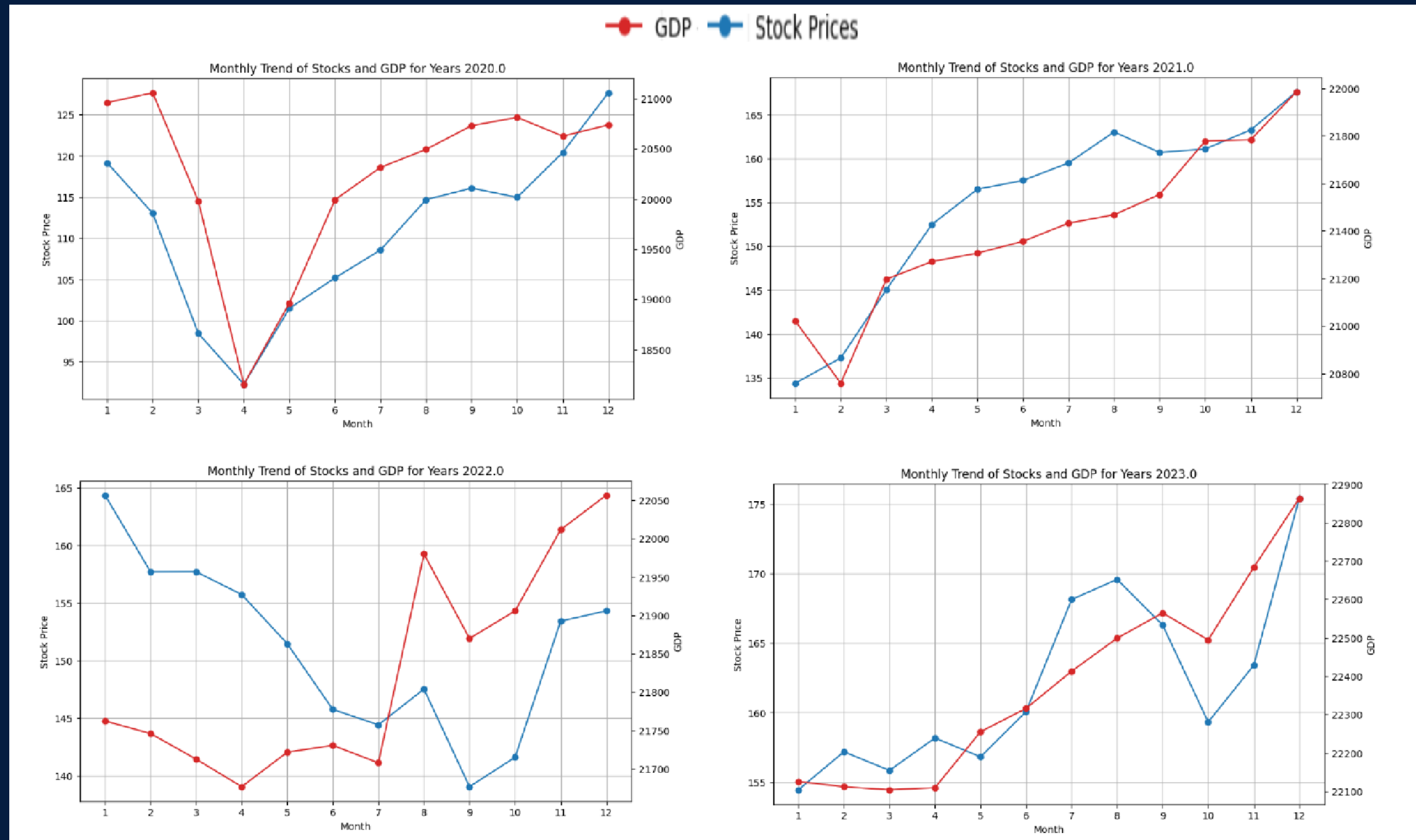
Unemployment Rate & GDP



- Reflect overall health and performance of an economy
- GDP & Unemployment Rate have an inverse relationship
- Correlations with stock market movements?
- Low unemployment rate and high GDP growth can boost investor confidence

GDP & Stock Price Relationship

- Monthly Data per Year
- Average Stock Price vs GDP
- Long-Term Trends
- Short-Term Volatility



Machine Learning Models

Training, Validation, Test Split

```
Train set shape: (1062, 51) (1062,)
Validation set shape: (354, 51) (354,)
Test set shape: (354, 51) (354,)
```

51 Features:

- Rank
- Sector
- Industry
- Monthly Stock Price (Open/Close)
- Monthly GDP
- Monthly Unemployment Rate

Feature Construction:

- Text Encoding
- Standard Scalar

Ridge Regression

Metric	Train	Val	Test
r2_score	0.9999997295127679	0.9999996960186028	0.9999995169662946
MSE	0.00040627745842896783	0.0004146064609040292	0.00042281839835754627

ElasticNetCV

Metric	Train	Val	Test
r2_score	0.9999749448619623	0.9999753345038238	0.9999756064535941
MSE	0.0040817237516502115	0.0040593932016837	0.003951960493191902

Linear SVR

Metric	Train	Val	Test
r2_score	0.9999028896660136	0.9999033604871286	0.9999015281382575
MSE	0.009851595448404936	0.009879195353520455	0.009847291911671288

What did work

- Visualizations for the relationships between the indicators
- Collecting data from 2020-2023 to predict the stock price for 2024 using API
- Find the trends between the economical indicators over stock price
- Predicting profit given stocks, and other financial indicators

What could be improved

- More historical data to capture long-term trends
- Integrate a column with real-time new and stock earning call data
- Utilize cross-validation techniques to improve model robustness



Thank You!