Stock Market Trends and Modeling

Matthew Shinder



The Stock Market



Sectors



Goals of this Project



Data Sources

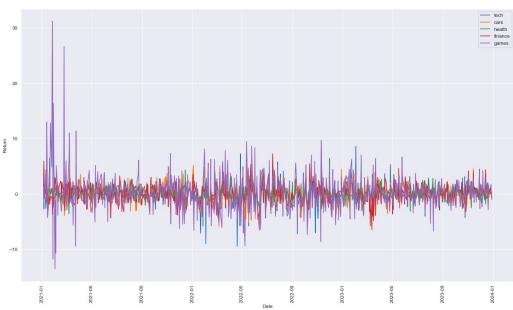


Data Wrangling / Cleaning

- No null values
- Created column 'returns'
 - Percentage Change Daily

EDA





games sector data features correlation

- 0.6

- 0.4

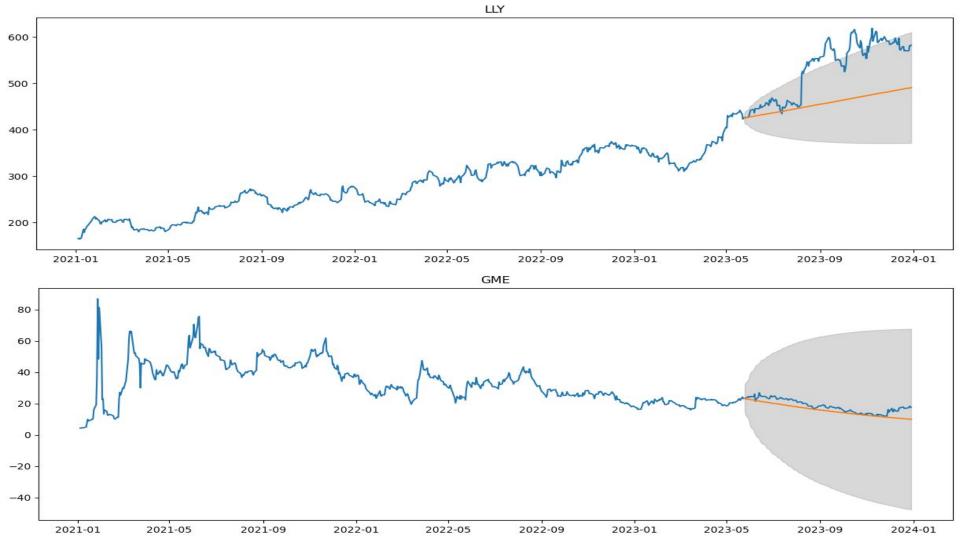
META	1.00	0.55	0.59	0.49	0.64	0.36	0.28	0.30	0.30	0.47	0.13	0.12	0.11	0.09	0.17	0.28	0.29	0.24	0.28	0.25	0.51	0.08	0.52	0.31	0.09
AAPL	0.55	1.00	0.61	0.46	0.66	0.55	0.38	0.39	0.42	0.55	0.18	0.24	0.26	0.31	0.23	0.33	0.35	0.32	0.37	0.26	0.59		0.63	0.37	0.05
AMZN	0.59	0.61	1.00	0.51	0.66	0.46	0.33	0.35	0.35	0.50	0.12	0.12	0.17	0.16	0.16	0.32	0.32	0.27	0.31	0.22	0.57	0.10	0.59	0.33	0.06
NFLX	0.49	0.46	0.51	1.00	0.45	0.38	0.30	0.30	0.22	0.38	0.10	0.09	0.08	0.13	0.13	0.23	0.24	0.18	0.24	0.16	0.43	0.04	0.46	0.28	0.02
GOOGL	0.64	0.66	0.66	0.45	1.00	0.42	0.35	0.37	0.37	0.51	0.15	0.18	0.18	0.23	0.19	0.33	0.34	0.31	0.34	0.24	0.59	0.06	0.61	0.35	0.07
TSLA	0.36	0.55	0.46	0.38	0.42	1.00	0.38	0.41	0.29	0.48	0.04	0.03	0.13	0.13	0.06	0.29	0.28	0.28	0.27	0.23	0.50	0.18	0.54	0.24	0.01
F	0.28	0.38	0.33	0.30	0.35	0.38	1.00	0.81	0.45	0.46	0.09	0.12	0.24	0.14	0.08	0.49	0.51	0.47	0.47	0.45	0.39	0.18	0.40	0.20	0.03
GM	0.30	0.39	0.35	0.30	0.37	0.41	0.81	1.00	0.47	0.48	0.11	0.14	0.26	0.16	0.10	0.58	0.57	0.55	0.55	0.51	0.39	0.17	0.40	0.23	0.06
TM	0.30	0.42	0.35	0.22	0.37	0.29	0.45	0.47	1.00	0.44	0.11	0.16	0.24	0.21	0.13	0.42	0.42	0.37	0.40	0.35	0.38	0.08	0.39	0.18	
RACE	0.47	0.55	0.50	0.38	0.51	0.48	0.46	0.48	0.44	1.00	0.17	0.13	0.16	0.22	0.19	0.37	0.37	0.33	0.40	0.29	0.51	0.14	0.55	0.32	0.09
PFE	0.13	0.18	0.12	0.10	0.15	0.04	0.09	0.11	0.11	0.17	1.00	0.39	0.30	0.27	0.30	0.16	0.16	0.15	0.19	0.18	0.11	-0.03	0.09	0.10	-0.02
JNJ	0.12	0.24	0.12	0.09	0.18	0.03	0.12	0.14	0.16	0.13	0.39	1.00	0.39	0.40	0.39	0.22	0.20	0.22	0.25	0.23	0.06	-0.00	0.05	0.17	-0.03
cvs	0.11	0.26	0.17	0.08	0.18	0.13	0.24	0.26	0.24	0.16	0.30	0.39	1.00	0.53	0.23	0.41	0.37	0.38	0.38	0.39	0.12	0.00	0.10	0.15	-0.04
UNH	0.09	0.31	0.16	0.13	0.23	0.13	0.14	0.16	0.21	0.22	0.27	0.40	0.53	1.00	0.34	0.28	0.23	0.26	0.31	0.23	0.17	-0.06	0.17	0.18	0.02
LLY	0.17	0.23	0.16	0.13	0.19	0.06	0.08	0.10	0.13	0.19	0.30	0.39	0.23	0.34	1.00	0.08	0.09	0.10	0.14	0.07	0.14	0.04	0.17	0.15	-0.00
BAC	0.28	0.33	0.32	0.23	0.33	0.29	0.49	0.58	0.42	0.37	0.16	0.22	0.41	0.28	0.08	1.00	0.81	0.82	0.84	0.75	0.32	0.09	0.31	0.21	0.06
С	0.29	0.35	0.32	0.24	0.34	0.28	0.51	0.57	0.42	0.37	0.16	0.20	0.37	0.23	0.09	0.81	1.00	0.76	0.78	0.69	0.35	0.13	0.33	0.21	0.02
WFC	0.24	0.32	0.27	0.18	0.31	0.28	0.47	0.55	0.37	0.33	0.15	0.22	0.38	0.26	0.10	0.82	0.76	1.00	0.75	0.72	0.28	0.08	0.29	0.15	0.05
JPM	0.28	0.37	0.31	0.24	0.34	0.27	0.47	0.55	0.40	0.40	0.19	0.25	0.38	0.31	0.14	0.84	0.78	0.75	1.00	0.67	0.32	0.07	0.34	0.22	0.03
USB	0.25	0.26	0.22	0.16	0.24	0.23	0.45	0.51	0.35	0.29	0.18	0.23	0.39	0.23	0.07	0.75	0.69	0.72	0.67	1.00	0.22	0.04	0.22	0.14	0.00
AMD	0.51	0.59	0.57	0.43	0.59	0.50	0.39	0.39	0.38	0.51	0.11	0.06	0.12	0.17	0.14	0.32	0.35	0.28	0.32	0.22	1.00	0.11	0.79	0.31	0.02
GME	0.08	0.12	0.10	0.04	0.06	0.18	0.18	0.17	0.08	0.14	-0.03	-0.00	0.00	-0.06	0.04	0.09	0.13	0.08	0.07	0.04	0.11	1.00	0.10	0.02	0.01
NVDA	0.52	0.63	0.59	0.46	0.61	0.54	0.40	0.40	0.39	0.55	0.09	0.05	0.10	0.17	0.17	0.31	0.33	0.29	0.34	0.22	0.79	0.10	1.00	0.31	0.04
EA	0.31	0.37	0.33	0.28	0.35	0.24	0.20	0.23	0.18	0.32	0.10	0.17	0.15	0.18	0.15	0.21	0.21	0.15	0.22	0.14	0.31	0.02	0.31	1.00	0.02
RBLX	0.09	0.05	0.06	0.02	0.07	0.01	0.03	0.06	0.15	0.09	-0.02	-0.03	-0.04	0.02	-0.00	0.06	0.02	0.05	0.03	0.00	0.02	0.01	0.04	0.02	1.00
	META	AAPL	AMZN	NFLX	19009	TSLA	ш	GM	MT	RACE	PFE	Ŋ	CVS	IN C	LLY	BAC	O	WFC	MAC	USB	AMD	GME	NVDA	A	RBLX

Modeling

- 1. ARIMA MODEL
 - a. Auto ARIMA function to automate process

Auto ARIMA

```
from statsmodels.tsa.arima.model import ARIMA
from pmdarima.arima import auto_arima
for stock in stockDataClose:
   chosen = stockDataClose[stock]
   chosen = chosen set_index('Date')
   train_data, test_data = chosen[3:int(len(chosen)*0.9)], chosen[int(len(chosen)*0.9):]
   model = auto_arima(train_data, start_p=1, start_q=1,
                     test='adf',
                     max p=5, max q=5,
                     m=1,
                     seasonal=True,
                     start P=0,
                     D=None,
                     trace=True.
                     error_action='ignore',
                     suppress_warnings=True,
                     stepwise=True)
   p,d,q = model.order
   prediction, confint = model.predict(n_periods=test_data.shape[0], return_conf_int=True)
   cf= pd DataFrame(confint)
   prediction_series = pd.Series(prediction.values,index=test_data.index)
   fig, ax = plt.subplots(1, 1, figsize=(15, 5))
   ax.plot(chosen['Close'])
   ax.plot(prediction series)
   ax set title(stock)
   ax.fill_between(prediction_series.index,
                   cf[1],color='grey',alpha=.3)
```



```
Stock Outlook
        15.24%
LLY:
NVDA:
        14.23%
GOOGL:
        7.14%
WFC:
        6.82%
UNH:
        6.81%
F:
        6.19%
AAPI:
        6.16%
RACE:
        5.21%
        3.68%
AMD:
JPM:
        1.74%
PFE:
        1.25%
        0.03%
JNJ:
CVS:
        -0.18%
BAC:
        -1.74%
META:
        -1.98%
EA:
        -2.02%
        -2.5%
TM:
        -6.37%
GM:
        -8.08%
TSLA:
        -8.88%
C:
        -9.11%
AMZN:
NFLX:
        -10.82%
USB:
        -12.41%
RBLX:
        -19.97%
        -57.42%
GME:
```

Tech Outlook: -1.72% Cars Outlook: -1.11% Health Outlook: 4.63% Finance Outlook: -2.9% Games Outlook: -12.3%

Analysis

- Best Outlook: Tech Stocks (LLY)
- Worst Outlook: Game Stocks (GME)
- Stocks not easy to predict
 - Otherwise, everyone would be rich
- ARIMA needs more tuning

Thank you!