# Time Series Forecasting of Stock Prices Across Various Sectors

- (Capstone Projects 2)

- Pallavi Bothra
- Data Science Career Track
- Mentor: A J Sanchez
-

# Stock Market Prediction



- Predicting the stock market is very challenging since it has far too many variables to be predicted.

- The market is inherently volatile and unpredictable.

- The traditional algorithms, like algorithmic trading do not fare much better than the average.

# Time Series Forecasting of Stock Price
## Capstone Project 2

- Total 1613 companies' stock data have been collected.

- Total 6 years stock data has been taken for each of the company (2012-2017).

- These companies are split into 10 sectors as follows:

-

- Communication Service Sector (70)
- Consumer Defensive Sector(180)
- Financial Sector (427)
- Industrial Sector (234)
- Technology Sector (194)

- Consumer Cyclical Sector (64)
- Materials Sector (132)
- Utility Sector (42)
- Energy Sector (104)
- Healthcare Sector (166)

# Time Series Forecasting of Stock Price:
# Data Wrangling

- Usual data cleaning procedure have been performed, e.g. replacing '$' with ' ', dropping columns etc.

- To find missing data we compare our dataset with New York Stock Exchange dataset and use interpolation method to fill the missing values (using web scraping).

- Extra features, such as companies' revenue, P/E ratios, interest rates have been extracted using web scraping.

- Data has been resampled into daily and quarterly frequency for modeling and exploring data, respectively.
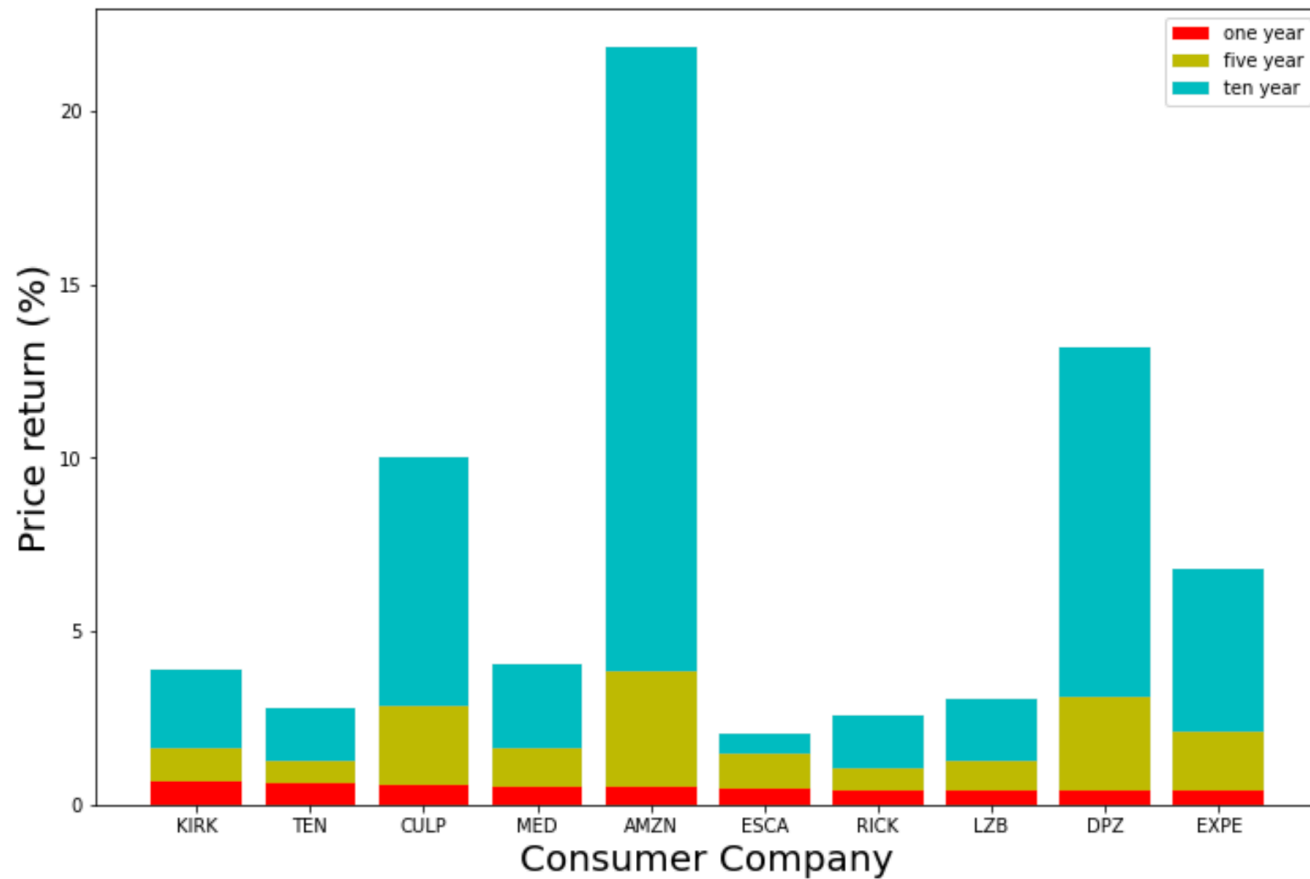
# Time Series Forecasting of Stock Price:
## Exploratory Data Analysis
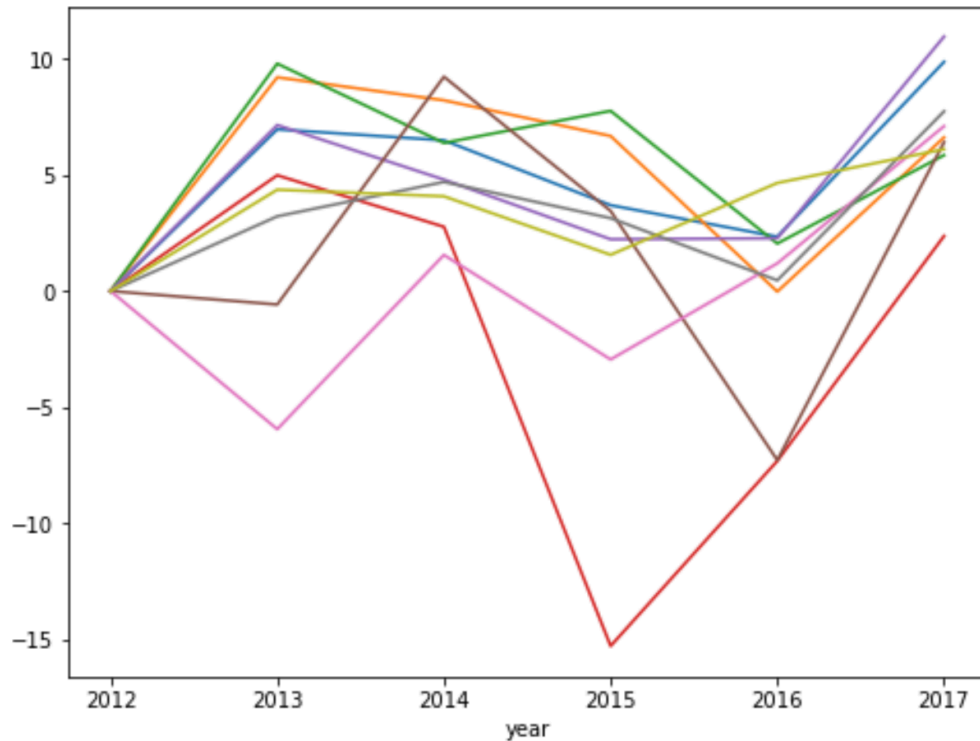


- Does the market follows a trend across sectors?

- Which sector is more volatile in 6 years?

- Which sector is more stable in 6 years?

# Time Series Forecasting of Stock Price:
## Exploratory Data Analysis



- To make data-driven informed decision, it is useful to check high momentum stock lists.

# Time Series Forecasting of Stock Price:
## Baseline Modelling (Manually)

- Usual statistical tests for time series modeling, such as KPSS test which checks for stationarity in the presence of a deterministic trend, and the ADF test for checking stationarity have been performed.

- Correlogram plots (ACF and PACF plots) have been plotted to detect the order for ARIMA(p,d,q) modeling.
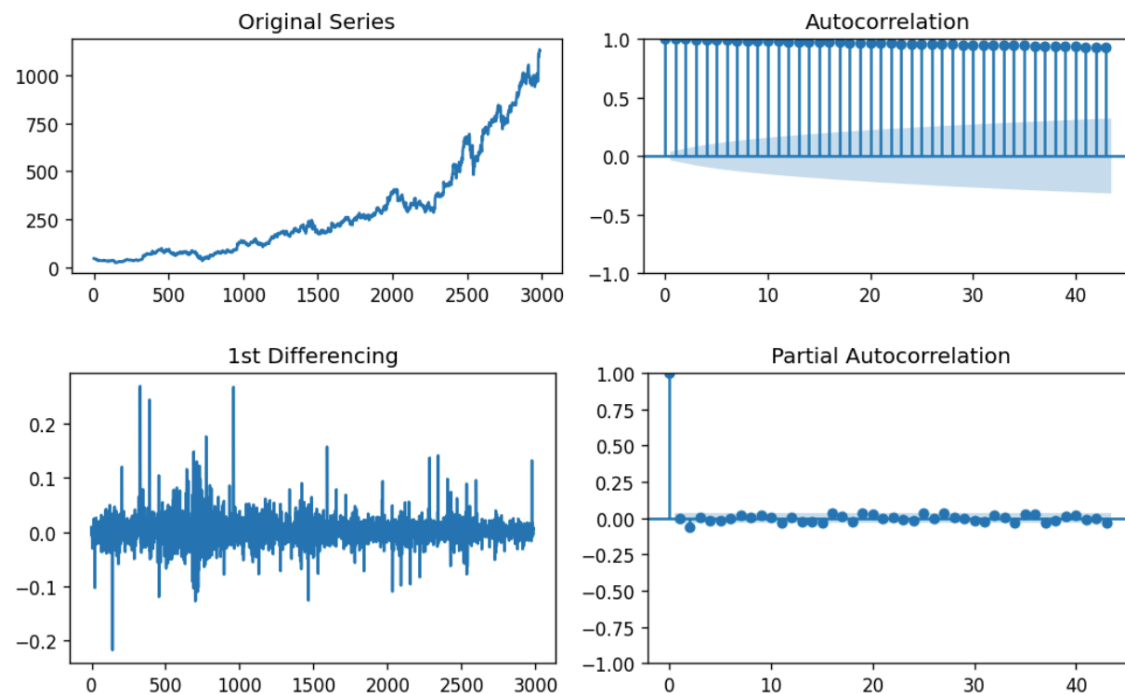
# Time Series Forecasting of Stock Price:
## ARIMA (p,d,q) Modelling

- ARIMA (0,1,0) modeling have been performed manually with only one company's dataset (Amazon)

- 30 days forecasting have been plotted with 95 % confidence interval.

- Residual vs Fitted value plot has been plotted to see if the residuals are random around the mean.

- Density distribution of error percentages for amazon stock price are shown, obtained from PMDARIMA and PROPHET model.



Forecast of AMZN Stock Price for next 30 days (ARIMA(0, 1, 0))



PE Density Plot
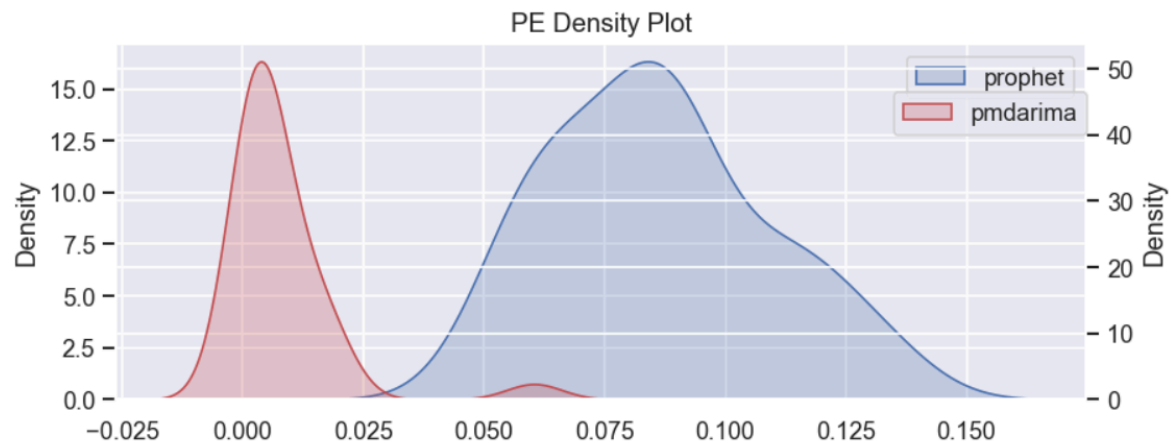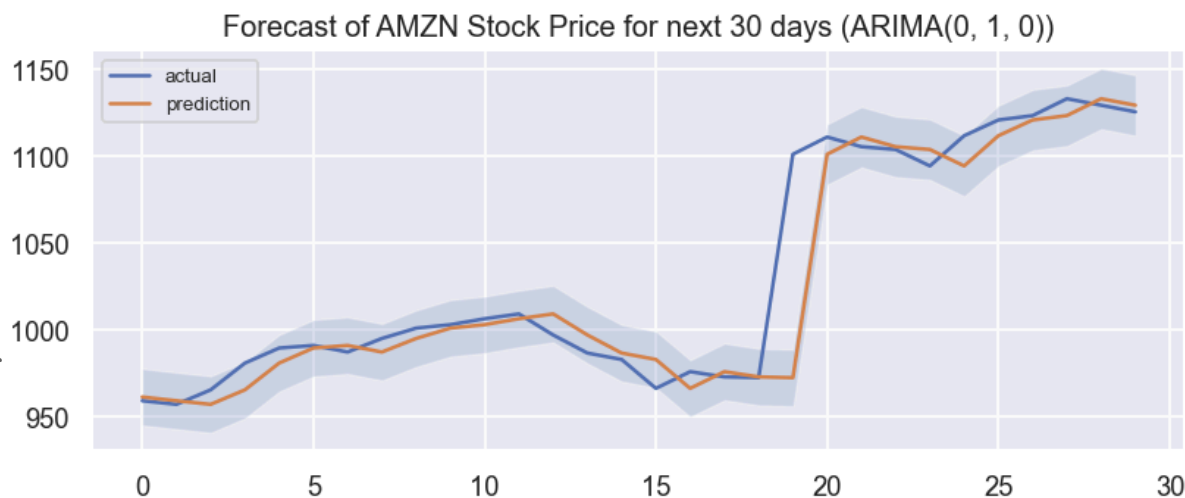
# Time Series Forecasting of Stock Price: Modelling

- Each of the 1619 company's 30 days stock price has been forecasted using PMDARIMA and FBPROPHET.

- The error metrics (MAPE, RMSE) have been computed for every forecasted results. (MAPE = Mean Absolute Percentage Error
  RMSE = Root Mean Squared Error)

|   | Company | PMDARIMA_RMSE | PMDARIMA_MAPE | Prophet_RMSE | Prophet_MAPE |
|---|---------|---------------|---------------|--------------|--------------|
| 0 | HD | 0.935167 | 0.004154 | 12.346563 | 0.074378 |
| 1 | CHDN | 1.498775 | 0.004915 | 30.911472 | 0.147477 |
| 2 | IHG | 0.381484 | 0.005273 | 5.553494 | 0.098873 |
| 3 | CCK | 0.435401 | 0.005526 | 4.383031 | 0.071684 |
| 4 | MCD | 1.253783 | 0.005906 | 15.886949 | 0.094600 |

- **Thank you for your attention**