

Capstone Project Presentation
Present

Stock Price Prediction

Present by
Jes Colbourne
Data Science & AI Program
Institute of Data 2022

CONTENT

	Page
Biography	3
Project context	4
Define	4
Design	5
Deliver	8
Summary	10
Appendix	

Biography

Jes Colbourne
Finance Officer



Education

- Bachelor's degree of Business Administration
- Advanced Diploma of Naturopathy
- Bachelor's degree of Business (Accounting)
- Data Science & AI Program

Professional experience

- Over 5 years of health analysis and consultation in retail environment
- Over 2 years of performance analysing and reporting of financial results

Project context

Australia's capitalisation of the listed equity market was about \$2.5 trillion dollars as of August 2021 which is quite small comparing to other larger overseas markets. However, our percentage of GDP is on par with UK and Japan who are in medium to high GDP rates. US has the highest growth where China and Euro area have lower rates.

35% of adult Australians invested on a securities exchange, 9 million of us are investors and 6.6 million of those hold shares or other no-exchange investments

COVID-19 pandemic creates a strong sense of uncertainty of our day-to-day life and our future. While many of brick-and-mortar businesses were brought down during the pandemic, online platform has seen its tremendous increase of the traffic.

One of the most popular platforms is YouTube where all kind of content and information are shared and consumed and almost half of younger investors (41%) educate themselves through YouTube videos!

Key challenges for these investors are underperformance stocks, hidden fees, market volatility and information overload that may lead to the loss of investment.

These challenges also make the problem interesting to find out how we can utilise data Science and AI tools to help minimise the risk and increase the chance of gaining returns on investment.

Define

Business aspects

The biggest question all investors want to know is if they can predict stocks prices and increase the chances of gaining more returns on investment?

The answer can value as high as their portfolio value and how much they are willing to risk but can be as low as the results from the risk itself.

To prevent losses and improve the chances of gaining we will utilise the power of data and artificial intelligence to help improve the business outcome.

Data science aspects

With the overwhelming availability of data, investors can get mixed information that can subsequently lead to misjudgement and making wrong decision when invest. So what information is relevant?

This project will look at historical stock prices that is easily access and available to everyone on yahoo finance website. We will also incorporate sentiment analysis score based on headline news dataset over a period of 19 years published by Australian Broadcasting Corporation available on Kaggle.com

As both data source is reliable and web based, investor will be able to progress to improve future project outcome by accessing real-time data when making prediction.

Design

Data exploration

Stock price data

Generally stock price is time series data which usually presents with trend and seasonality over a period of time, so we have to remove them from the dataset to make it stationary which means. However, for this exercise we will predict the movement of stock prices by observing the difference of closing and opening stock price which also making the data more stationary.

This project is looking at Aztrazeneca stock prices (AZN) over the past 19 years which will cover the same period of the headline news data set as well as covering the trend before and after the effect of COVID-19 pandemic.

Headline news data

There can be several headline news each day and they can be good news and bad news. To observe sentiment effect from the news, we use the mean of polarity compound score which is a normalized score of the effect of the news each day so that we can compare the overall daily score to the movement of stock price on that day.

Data analysis

Based on observation of the trend of AZN price movement and sentiment score over 19 years period, AZN prices seem to be affected by major events of diseases outbreak such as SARS in 2003, Swine flu in 2010, MERS in 2015 and COVID-19 in 2019. Although sentiment score does not seem to get direct effect at the peak of each event, the scores fall on the negative value for a prolong period after each event as seen in figure1 and 2 below.

When looking at correlation between 2 features the heatmap shows high correlation among stock price features but low correlation of price and sentiment score as shown in figure 3.

Figure 1

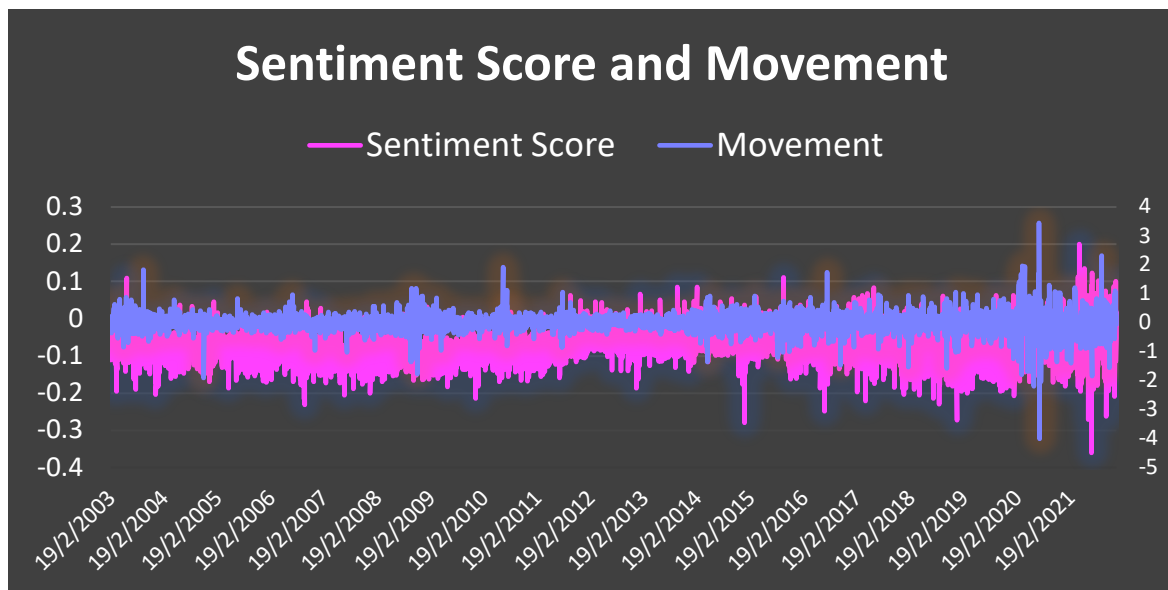


Figure 2

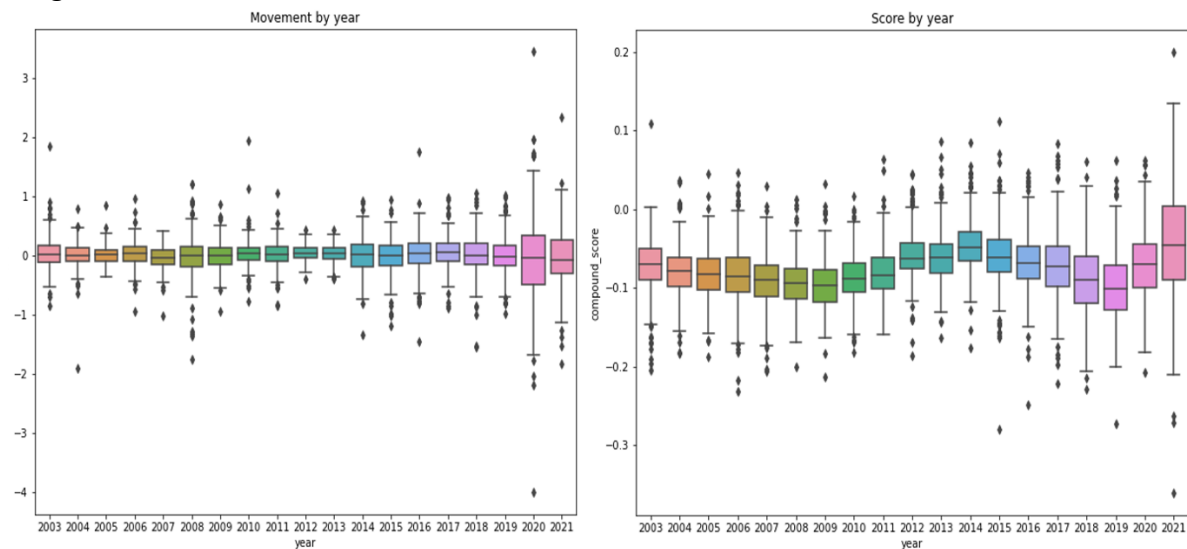
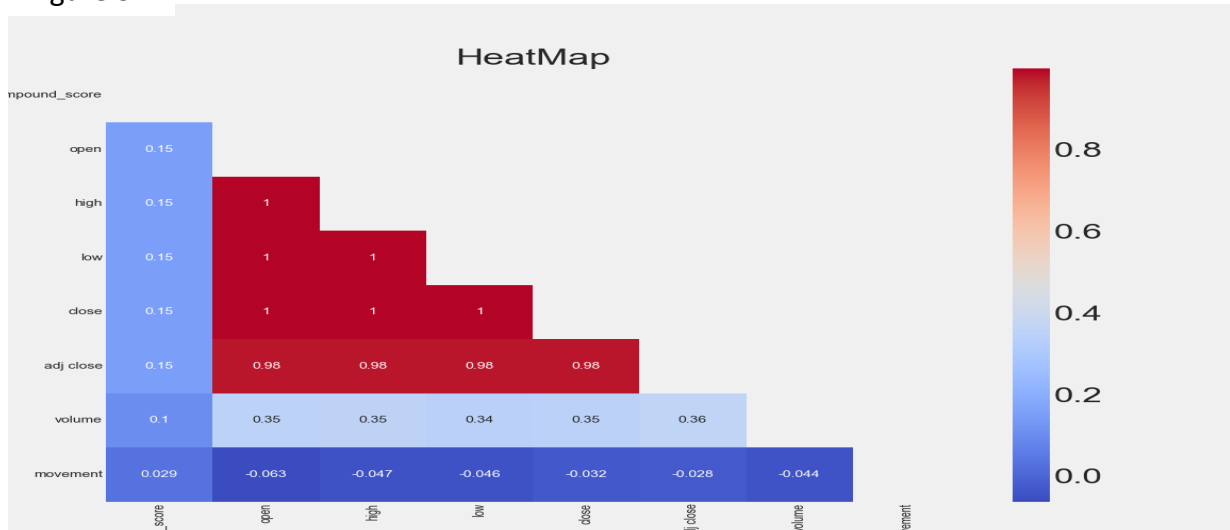
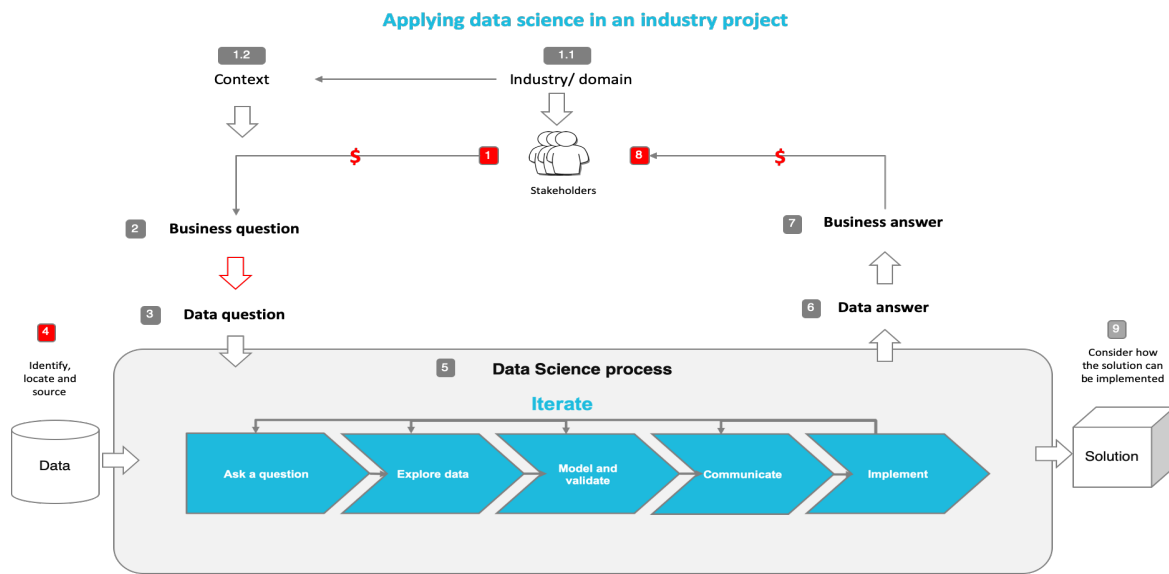


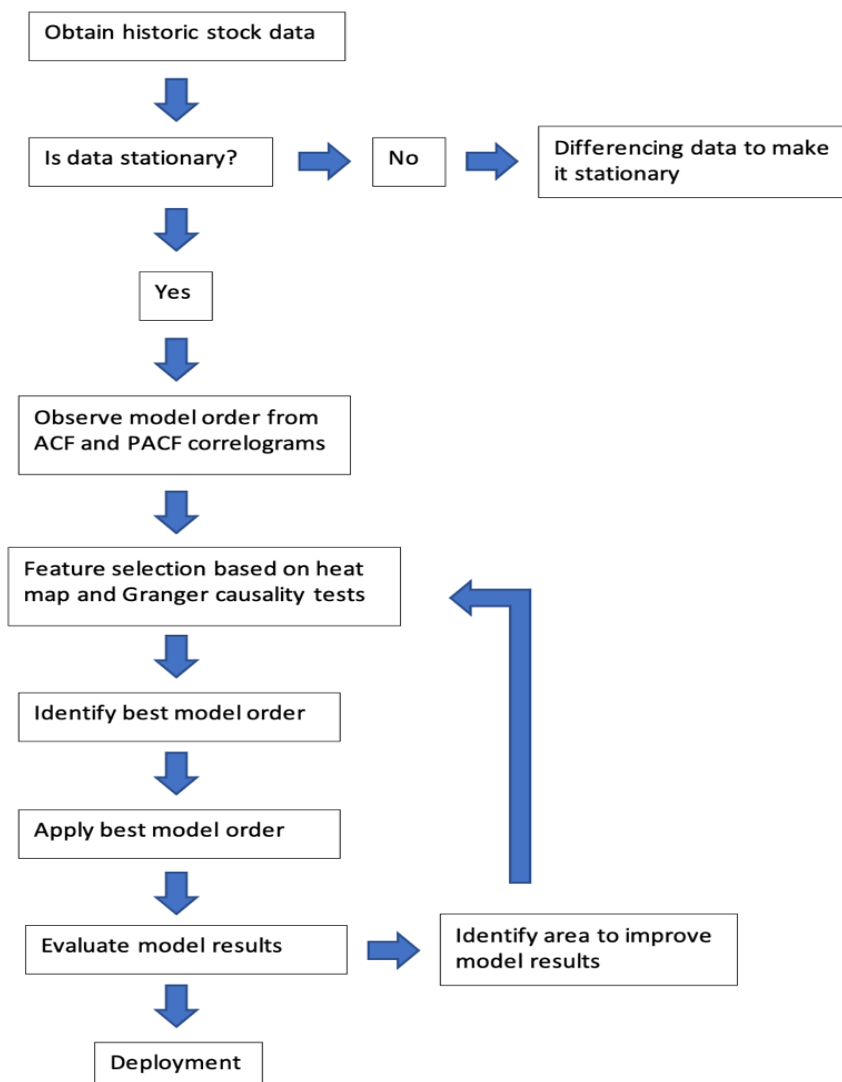
Figure 3



Overall process flow



Modelling process flow

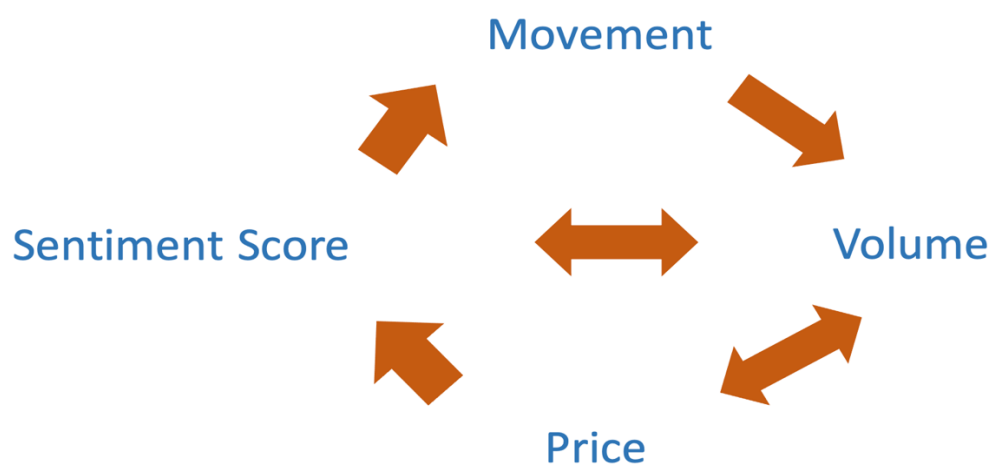
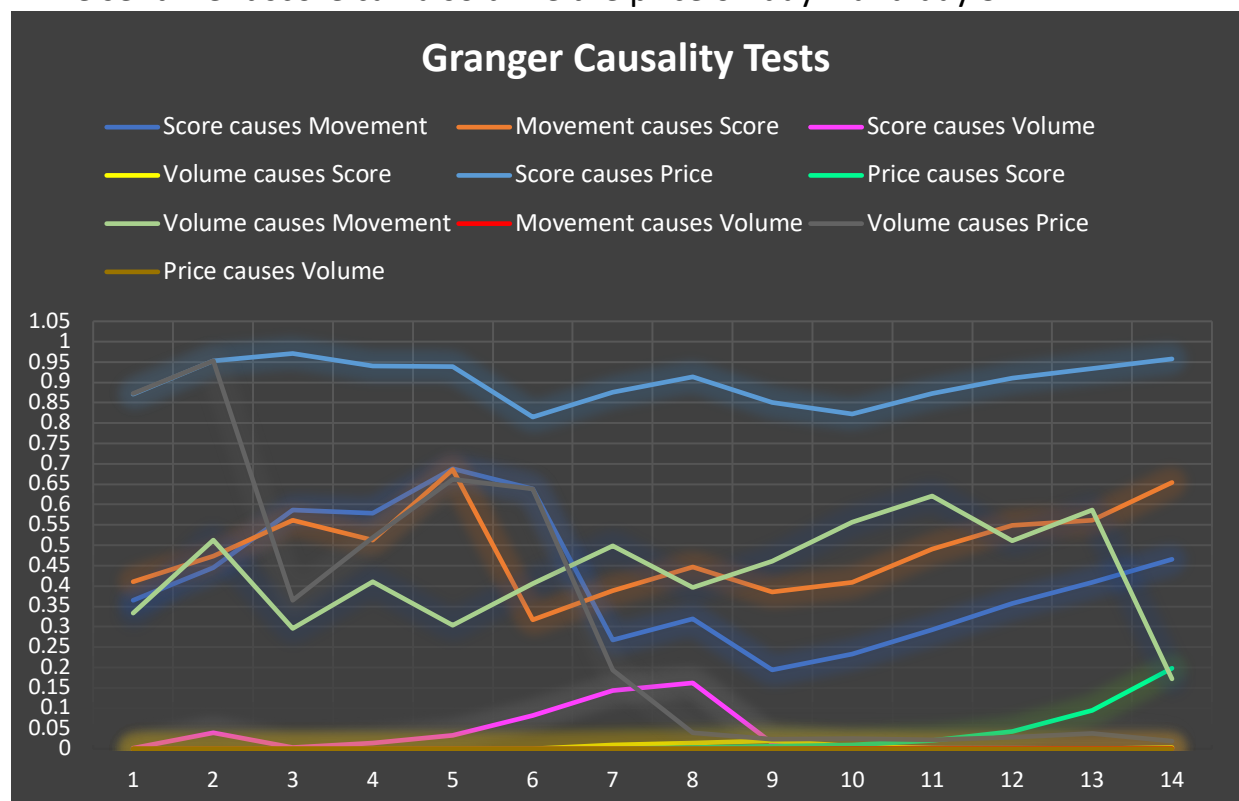


Deliver

Feature engineering

The most important features in this project are movement of stock price and sentiment score. The reason that these two are chosen is to provide validation whether stock price movement can be predicted based on sentiment score.

Granger Causality test below shows p-value results of each feature pair. Stock volume and price can drive sentiment score results on the lag of 1-6 days while sentiment score can also drive the price on day 1 and day 3.



Machine model

This project is based on Autoregressive Integrative Moving Average model (ARIMA) because it considers of time series type of data that predict future value based on past value and it is an appropriate and simple model for prediction when using fewer features.

A simple model like ARIMA can be very easy to apply by most users, effective, and low cost of maintenance so investors can maximise the total gain of their investment returns.

We also use auto.arima function to help perform a search for the best possible order of the model then we use this order combination to fit onto dataset to predict the results

Modelling

Predict movement with sentiment score

Features: Sentiment score and price movement (sentiment score is being used as an exogenous feature to price movement)

Best model order: (2, 1, 2)

Total fit time on training set: 164.484 seconds

Model performance metrics: Root mean square error

Model results: rmse 0.121 on test set and 0.5649 on unseen set

Predict price movement next day base on the actual movement

Features: Price movement

Best model order: (3, 1, 1)

Total fit time on training set: 64.040 seconds

Model performance metrics: Root mean square error

Model results: rmse 0.121 on test set and 0.5425 on unseen set

Model selection

Based on rmse result of 0.543 on unseen set when predicting just on price movement, we will choose to deploy ARIMA model without sentiment score because it may rather just create white noise to the model than enhancing model results

Deployment

We deploy ARIMA(3, 1, 1) as a base model in a simple simulation environment to random buy and sell stocks with starting investment value of \$1,000 from 2003. By the end of 2021, the model made a total return of 207.96% with total amount value of \$3,079.59 as AZN stock price jumped from less than \$10 per share to over \$60 per share over 19 years period. However,

when model is set to buy only if last return was positive, it made a loss of 36.31% with \$363.12 loss.

On the other hand, when apply this model on Qantas stocks, the model made a loss of 19.38% or \$193.82 total in value and made a gain of 34.86% when only buy if last return was positive.

Summary

Stock price movement prediction is possible when consider other factors that may have direct and indirect impact to the chosen stock based on its nature, industry and economic impacts around it.

Key takeaway from this project

1. Invest time in data analysis to gain more domain knowledge of chosen stock, its market and overall economy
2. Analysing all features that may have direct and indirect impact on model results but keep the model simple to save on training time and maintenance
3. Invest time in training and tuning the model for better results

Next steps

1. Study technical analysis of market trend
2. Set up appropriate strategy of your investment plan
3. Test and set up rules-based strategy that is profitable
4. Optimise strategy outcome ie stop loss, buying target, risk-reward ratio, and time when the strategy work best
5. Risk management plan - managing the loss, keep you out of emotion when trade
6. Practice simulation/demo trading before investing real money
7. Discipline - don't gamble and stick to the strategy

Appendix

1. <https://www.rba.gov.au/publications/submissions/financial-sector/background-on-the-australian-listed-equity-market-2021-09/index.html>
2. <https://www2.asx.com.au/blog/australian-investor-study>
3. <https://github.com/ritvikmath/YouTubeVideoCode/blob/main/ARMA%20Stock%20Forecasting.ipynb>
4. <https://medium.datadriveninvestor.com/time-series-prediction-using-sarimax-a6604f258c56>
5. https://www.statsmodels.org/dev/examples/notebooks/generated/statespace_sarimax_faq.html
6. <https://github.com/nachi-hebbar/Multivariate-Time-Series-Forecasting>
7. <https://www.kaggle.com/datasets/therohk/million-headlines?select=abcnews-date-text.csv>
8. <https://finance.yahoo.com/quote/AZN/>
9. <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>
10. <https://www.frontiersin.org/articles/10.3389/fmicb.2020.631736/full>