Sprint 2

Stock data and Random Matrix Theory

Hugo C Marrochio

S&P Stock Data Predictions

Goals

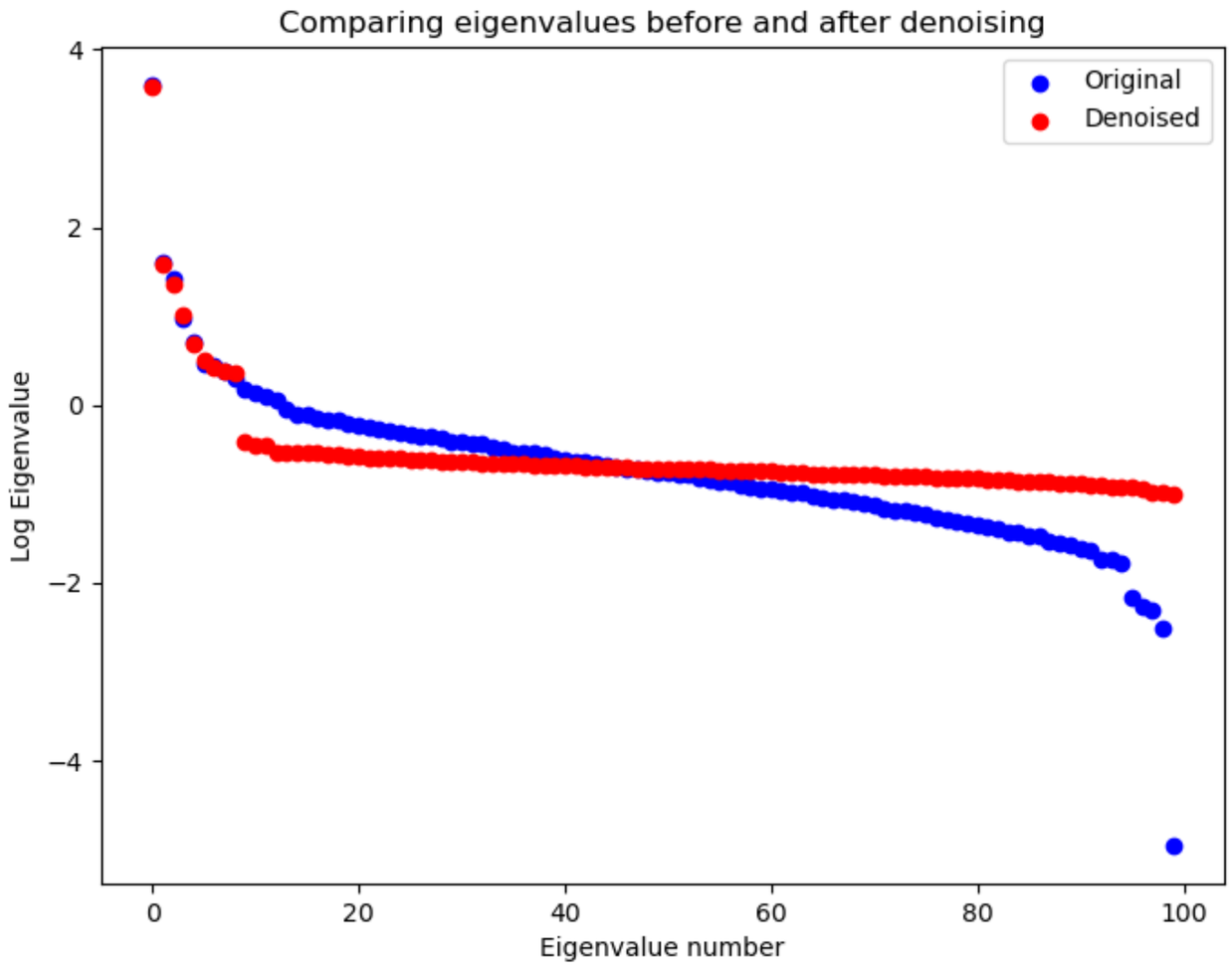
- Use ideas from Physics and Mathematics about Randomness to investigate stock movements.
- Practice different Machine Learning techniques.
- Characterize clusters, predict future prices, build trading portfolios.

Plan for the Presentation

- Explain denoising techniques
- Show results of clustering algorithms
- Show preliminary results for stock price forecasting

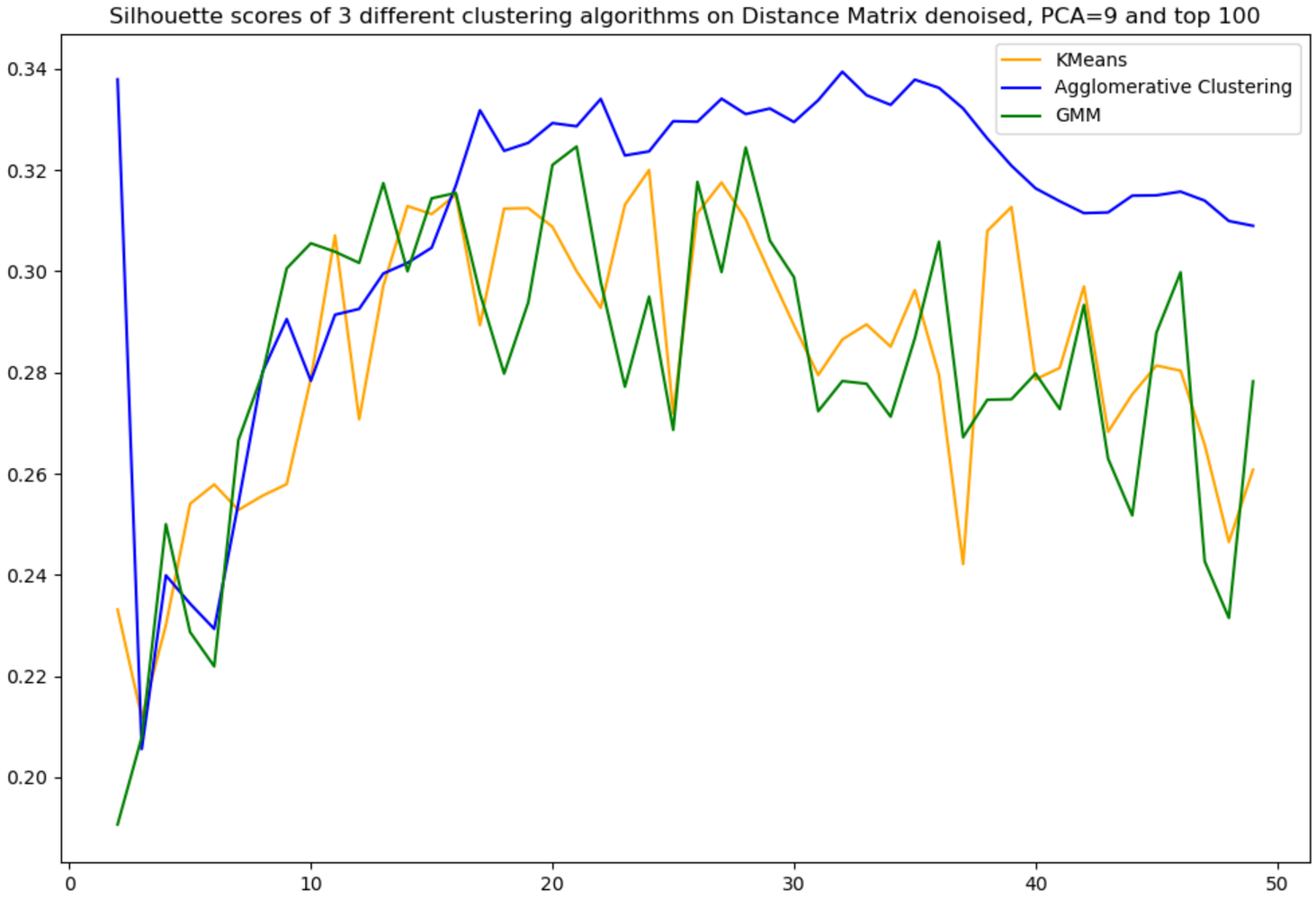
Eigenvalue Denoising

- Think of eigenvalues as the directions of a PCA decomposition
- Denoising simply “flattens” the spectrum in the random range
- Since there is less structure due to randomness, clustering algorithms tend to perform better
- Eigenvalues outside random range should inform PCA components.

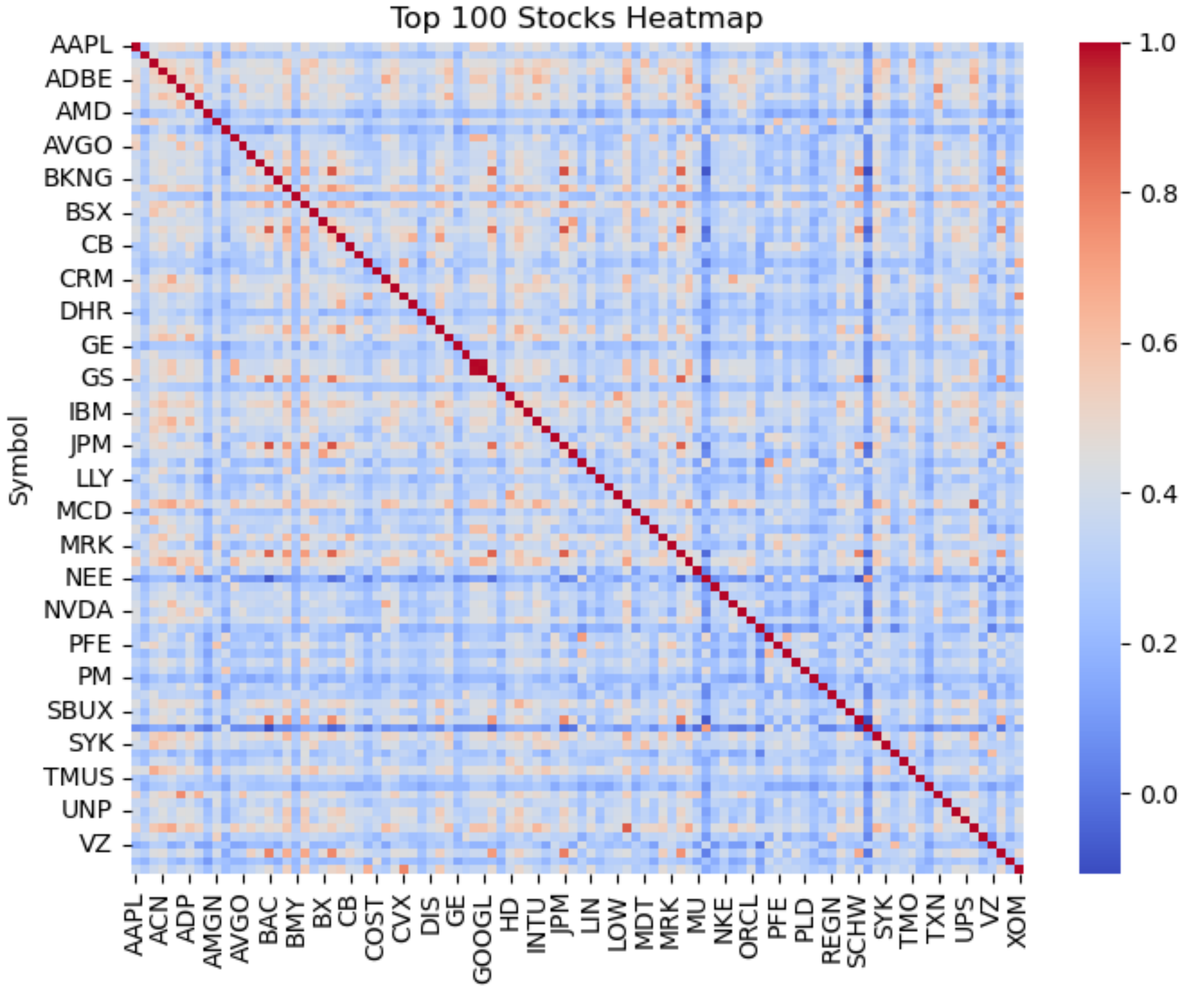


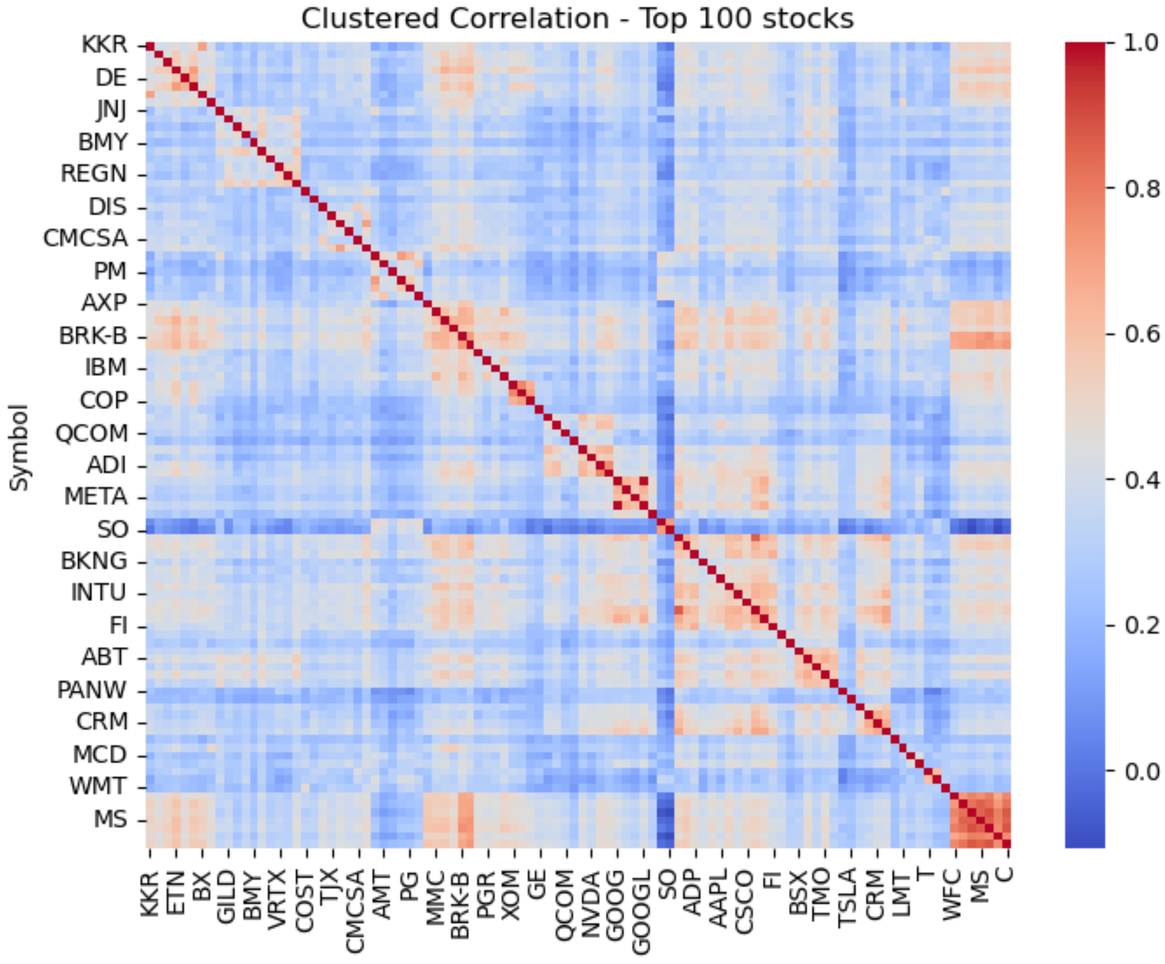
Clustering - Silhouette score

- Silhouette score in order to find algorithm and cluster size
- Used denoised correlation matrices as input for clustering
- For interpretability and high silhouette score, chose Agglomerative Clustering and k=16.



Clustered Correlation Matrix Comparison





Local Cluster Structures!

Was denoising useful?

- Was denoising useful for clustering? Results so far promising!
- Denoised correlation matrix created more uniformly distributed clusters.
- Example: Results from original data contain a cluster only with Google stocks!

Denoised

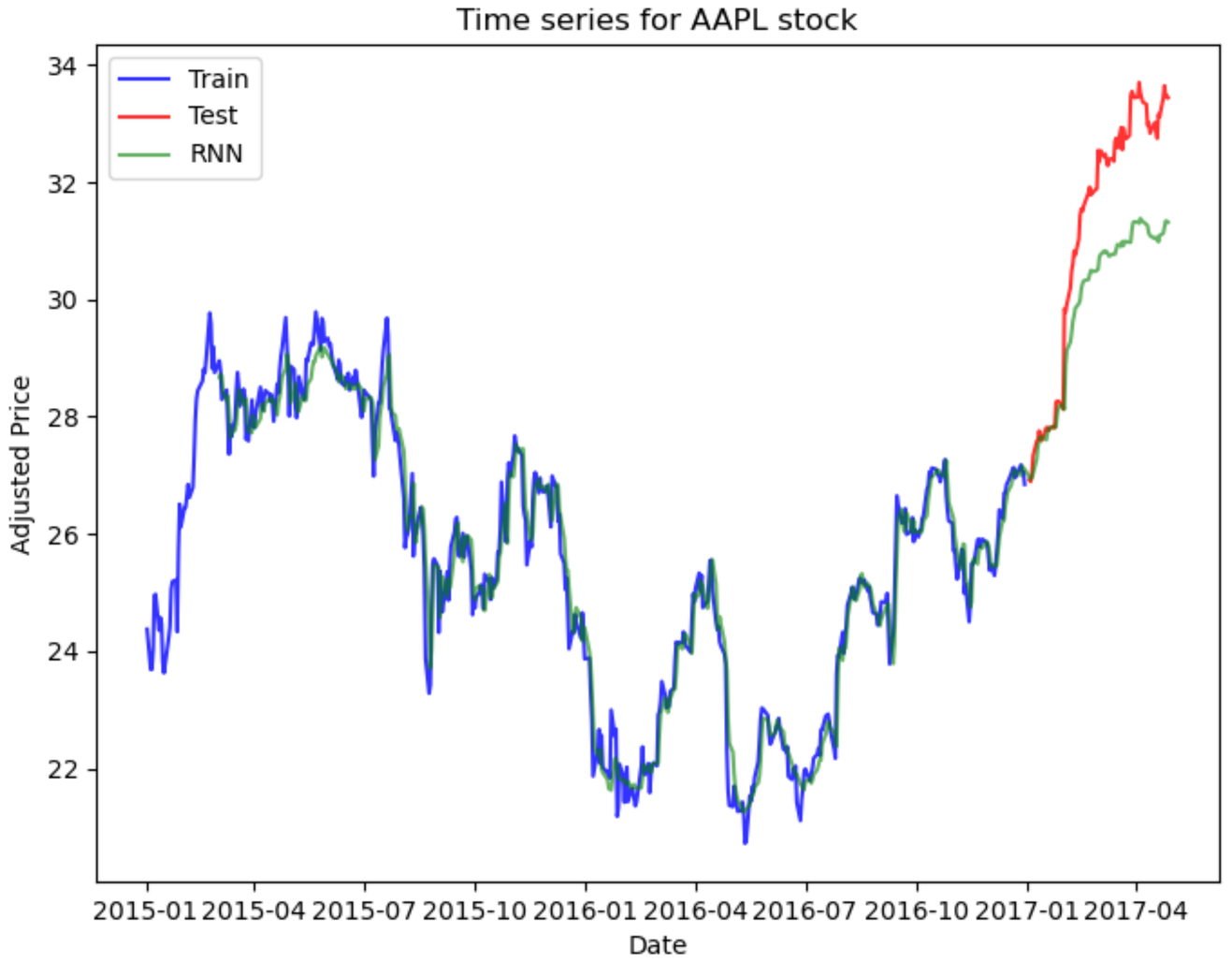
	Exchange	Symbol	Shortname	Longname	Sector	Industry
3	NMS	GOOG	Alphabet Inc.	Alphabet Inc.	Communication Services	Internet Content & Information
4	NMS	GOOGL	Alphabet Inc.	Alphabet Inc.	Communication Services	Internet Content & Information
5	NMS	AMZN	Amazon.com, Inc.	Amazon.com, Inc.	Consumer Cyclical	Internet Retail
6	NMS	META	Meta Platforms, Inc.	Meta Platforms, Inc.	Communication Services	Internet Content & Information
22	NMS	NFLX	Netflix, Inc.	Netflix, Inc.	Communication Services	Entertainment

Original

	Exchange	Symbol	Shortname	Longname	Sector	Industry
3	NMS	GOOG	Alphabet Inc.	Alphabet Inc.	Communication Services	Internet Content & Information
4	NMS	GOOGL	Alphabet Inc.	Alphabet Inc.	Communication Services	Internet Content & Information

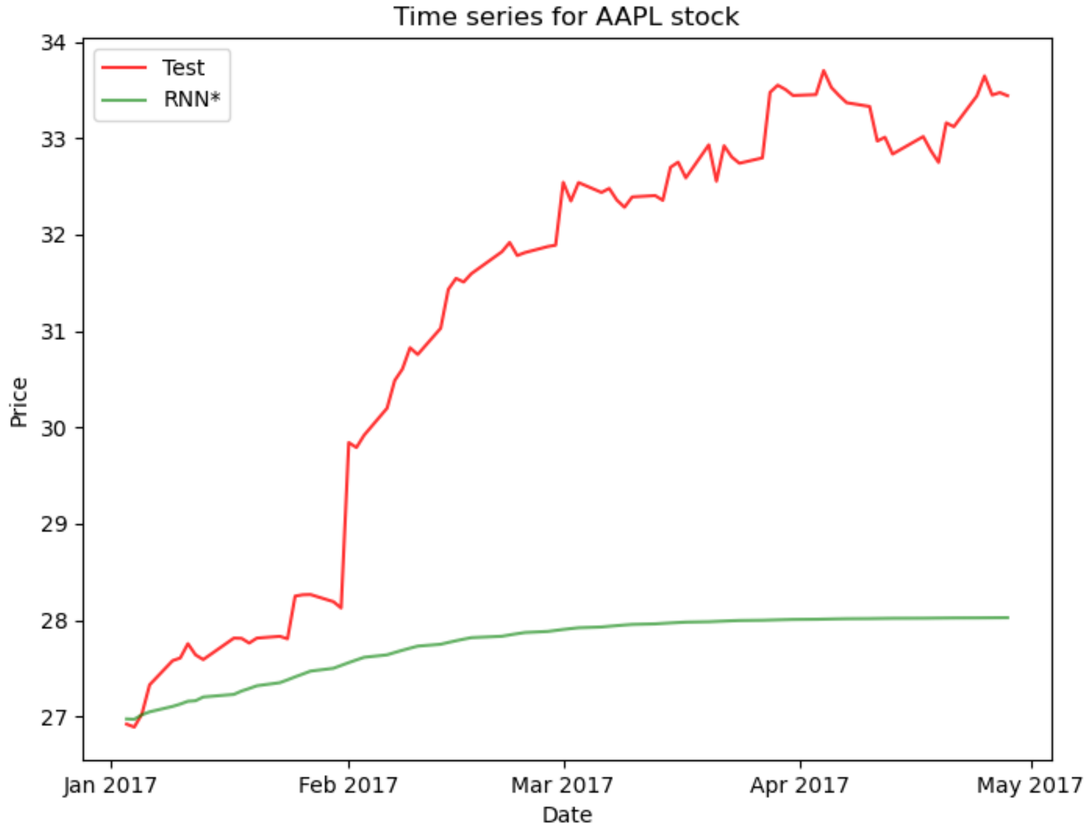
Preliminary - Time Series Price Prediction

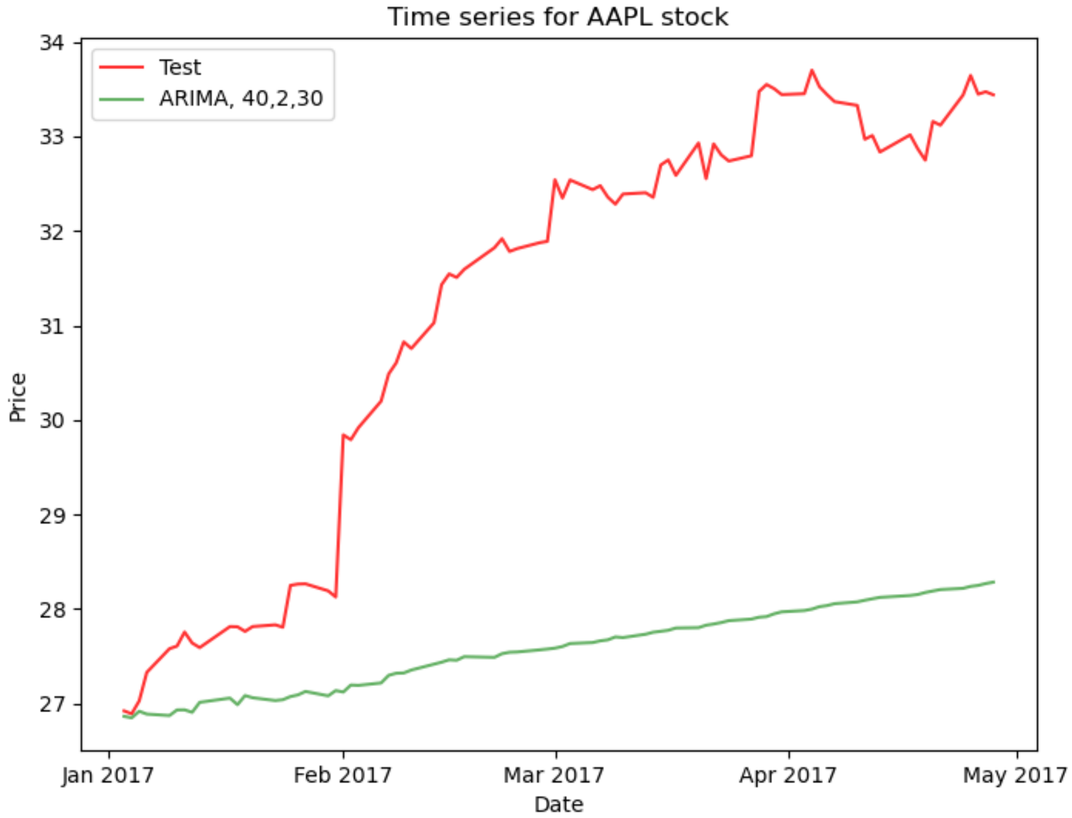
- First, the good news!
- Simple RNN architecture achieved good agreement with test data, even if test data was an atypical high compared to training.
- Caveat: we only train with training data, but we forecast with all the data up to the previous date.



Preliminary - Time Series Price Prediction

- If we use only the test data to forecast, so all the future points are outputs of the model, prediction is much less precise.
- I called this forecasting result RNN* (also showing similar forecasting technique for ARIMA, with similar results)





Preliminary - Time Series Forecast

- Next, I plan to complete this table with predictions from different RNN architecture structures.
- I also want to understand the downsides of forecasting far into the future, and how this can affect investing strategies.
- After this forecasting table is done, I want to move forward to portfolio optimization!

Model	Parameters	MSE
Naive Avg	Training Data Average	0.356
ARIMA	(40,2,30)	0.202
RNN*	ReLU, avoid overfitting	0.208

Model	Learning Rate (ADAM)	Activation	MSE
RNN	0.005	'ReLU'	0.028
GRU			
LSTM			
Transformer			