

tcs

IBM

Infosys

accenture

# STOCK PREDICTIONS NSE AND S&P 500

## DATA NINJA 7:

Akshay Deshmukh  
Jai Trivedi  
Nilabh Chaturvedy  
Niyathi Aiella  
Tanveer Singh Sapra

# Problem Statement

The financial-investment market is a highly volatile and sensitive market which attracts many investors. These investors are looking for a reliable way to know the risks and returns associated with a stock listing.

We aim to predict the stock price and volume of 4 International Tech firms, namely: **Accenture, Infosys, IBM, and TCS**. We have implemented predictive and forecasting analysis to study the price variation of the stocks and predict their trends.

By predicting the price of the stocks, our clients are more informed of their investment decisions and have a clearer picture of the prospects of their investments.

Our data analysis revealed the trends in each of the stocks. We were also able to compare our predictive model with the actual results and retrain our model on regularly streamed data.

By studying the variability in the stock prices, we are also able to recommend stocks that are low and high risk.

# Pipeline Formulation

## Data flow named stocksjob\_new3

-- Pumps data into bigquery and storage

stocksjob\_new3

CLONE

STOP

CREATE SNAPSHOT

IMPORT AS PIPELINE

SHARE

JOB GRAPH

EXECUTION DETAILS

JOB METRICS

RECOMMENDATIONS

AUTO-SCALING

Job steps view

Graph view

CLEAR SELECTION

ReadPubSubTopic

Running

1 sec

1 stage

ConvertMess...ToTableRow

Running

2 sec

1 stage

WriteSuccessfulRecords

Running

30 min 57 sec

2 stages

Flatten

0 sec

0 stages

WrapInsertionErrors

Running

0 sec

1 stage

WriteFailedRecords

Running

0 sec

2 stages

WriteFailedRecords2

Running

0 sec

2 stages

Job info

Job name

stocksjob\_new3

Job ID

2022-12-12\_21\_53\_45-11275378726300055526

Job type

Streaming

Job status

Running

SDK version

Apache Beam SDK for Java 2.43.0

Job region

us-central1

Worker location

us-central1

Current workers

1

Latest worker status

Auto-scaling: Reduced the number of workers to 1 based on low average worker CPU utilisation and the pipeline having sufficiently low backlog and keeping up with input rate.

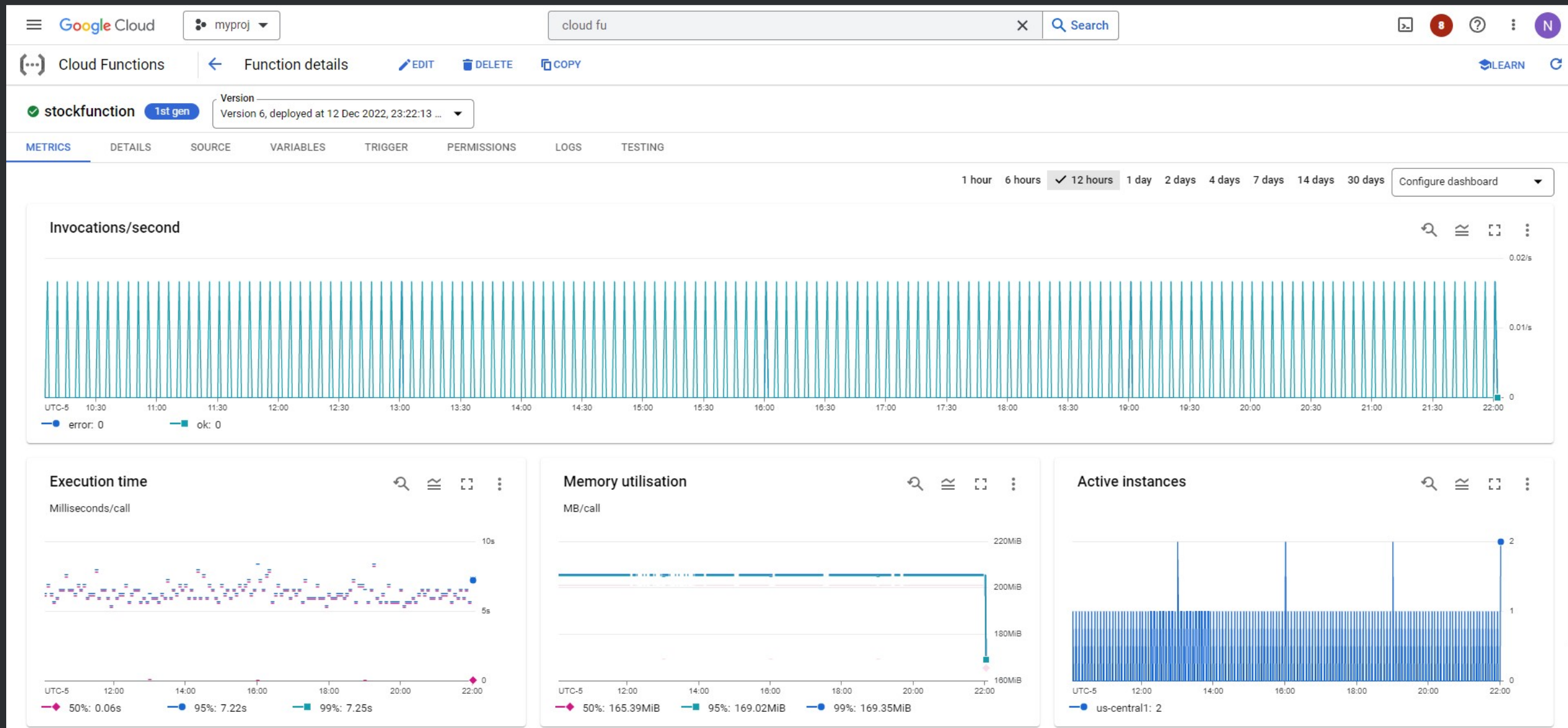
Start time

13 December 2022 at

# Pipeline Formulation

## Cloud Function named stockfunction:

- This inturn gets triggered by cloud Scheduler which runs every 5 minutes
- Pulls data from YahooFinance API from the code



# Pipeline Formulation

## Cloud scheduler was enabled:

-- This scheduler runs every 5 minutes enabling the cloud function to run and pull the data

☐ [stock\\_streaming\\_scheduler](#)  Success us-central1 Enabled \*/5 \* \* \* \* (America/Indianapolis)


## Cloud storage is populated with the data:

-- All the data that is coming from the archive and the yahooFinance is going into cloud storage bucket


**mybucket0447**






Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Not public	None

[OBJECTS](#) [CONFIGURATION](#) [PERMISSION](#) [PROTECTION](#) [LIFECYCLE](#) [OBSERVABILITY](#) [NEW](#)

Buckets > mybucket0447 > stocks3 > timestamp=2022-12-13T06:41:28 

[UPLOAD FILES](#) [UPLOAD FOLDER](#) [CREATE FOLDER](#) [TRANSFER DATA](#) [MANAGE HOLDS](#) [DOWNLOAD](#) [DELETE](#)

Filter by name prefix only  Filter Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created 	Storage class	Last modified
<input type="checkbox"/>	 <a href="#">symbol=ACN/</a>	—	Folder	—	—	—
<input type="checkbox"/>	 <a href="#">symbol=IBM/</a>	—	Folder	—	—	—
<input type="checkbox"/>	 <a href="#">symbol=INFY.NS/</a>	—	Folder	—	—	—
<input type="checkbox"/>	 <a href="#">symbol=TCS.NS/</a>	—	Folder	—	—	—

# Historical Analysis

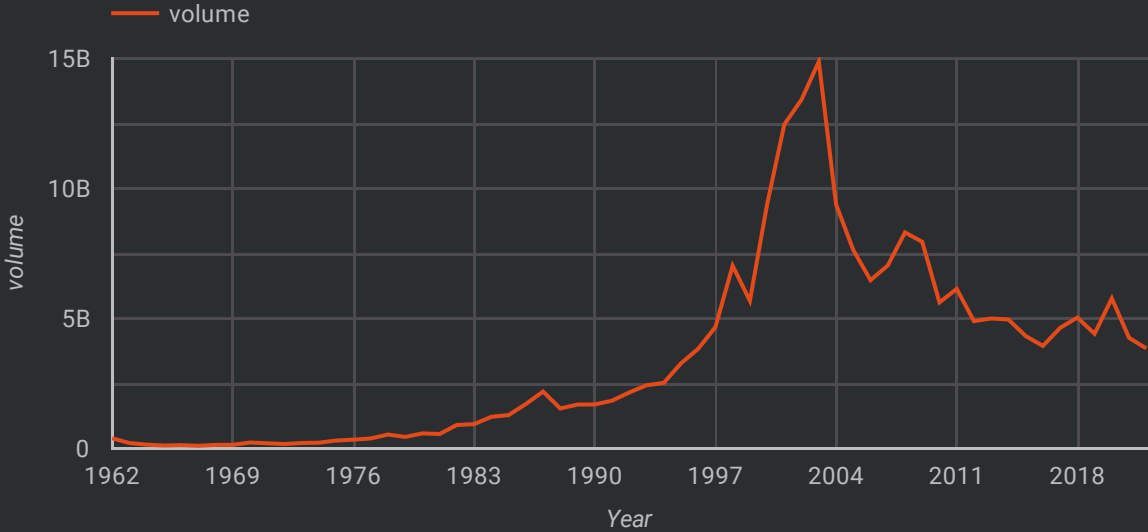


Average High

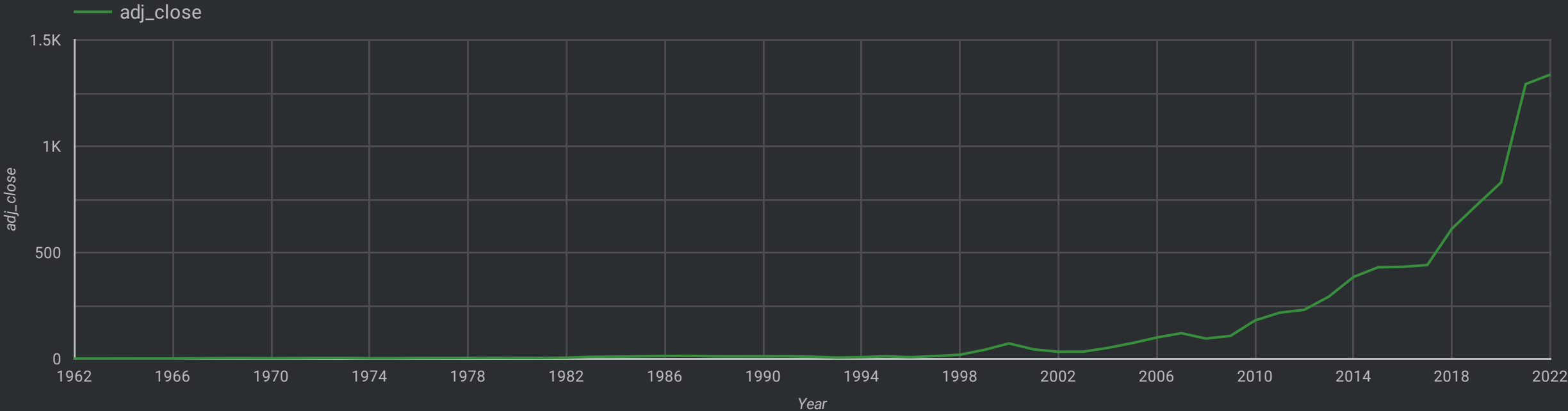
	symbol	high ▾
1.	TCS	1,055.7
2.	INFY	375.52
3.	ACN	96.34
4.	IBM	61.12

1 - 4 / 4 < >

Volume Traded



Adjusted Close



# Evaluation Of Models

## ARIMA(Autoregressive Integrated Moving Average)

DETAILS	TRAINING	EVALUATION	SCHEMA					
Time series ID	Non-seasonal P <span>?</span>	Non-seasonal D <span>?</span>	Non-seasonal Q <span>?</span>	Has drift	Log likelihood	AIC	Variance	Seasonal period
ACN	1	2	4	False	-9,341.17	18,694.341	0.643	Weekly, Yearly
IBM	0	1	5	False	-12,837.109	25,686.219	0.186	Weekly, Yearly
INFY	0	2	5	False	-25,746.824	51,505.649	11.076	Weekly, Yearly
TCS	0	2	5	False	-26,681.341	53,374.681	78.789	Weekly, Yearly

## REGRESSION - Volume Prediction

volume_prediction_boosting			
DETAILS	TRAINING	EVALUATION	SCHEMA
Mean absolute error	2,489,842.0023		
Mean squared error	43,586,161,573,481.52		
Mean squared log error	2.6431		
Median absolute error	1,053,025.5		
R squared	0.5543		

Boosting Regression

volume_prediction_dnn			
DETAILS	TRAINING	EVALUATION	SCHEMA
Mean absolute error	6,630,523.1794		
Mean squared error	141,764,249,078,145.53		
Mean squared log error	122.659		
Median absolute error	4,019,247.9788		
R squared	-0.4495		

DNN Regressor


volume\_prediction\_rf

DETAILS

TRAINING

EVALUATION

SCHEMA

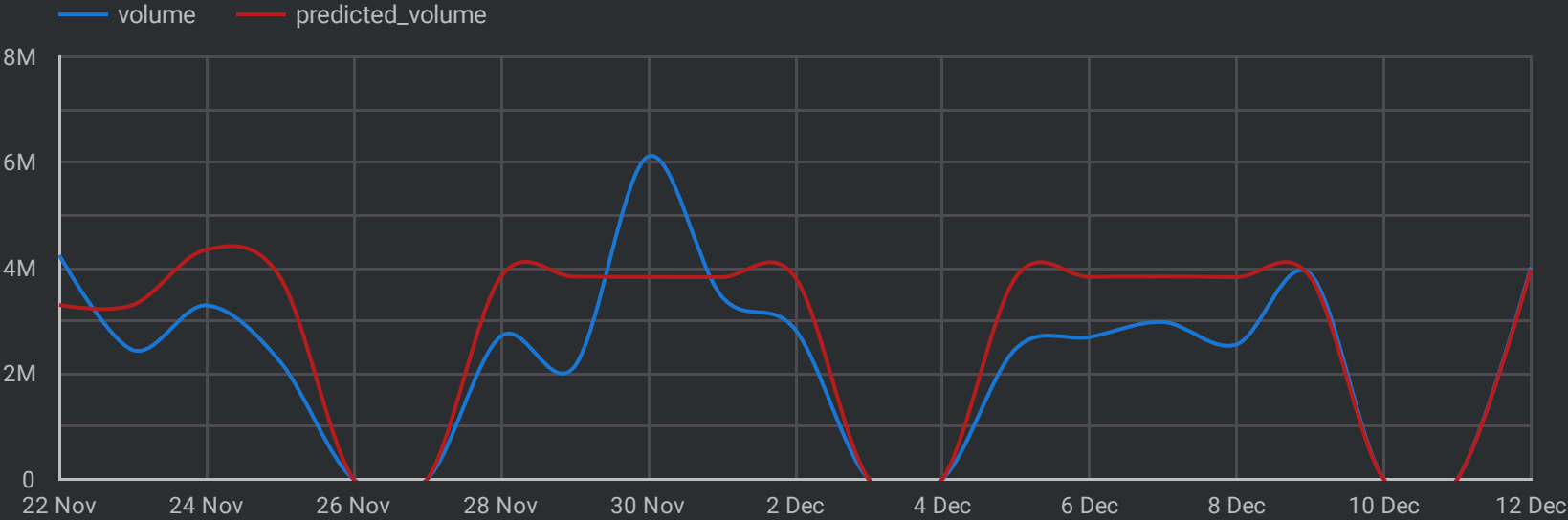
 This model has no evaluation metrics.

Mean absolute error	2,361,132.5923
Mean squared error	32,116,559,093,516.113
Mean squared log error	1.7534
Median absolute error	1,027,287.5
R squared	0.6063

Random Forest Regressor

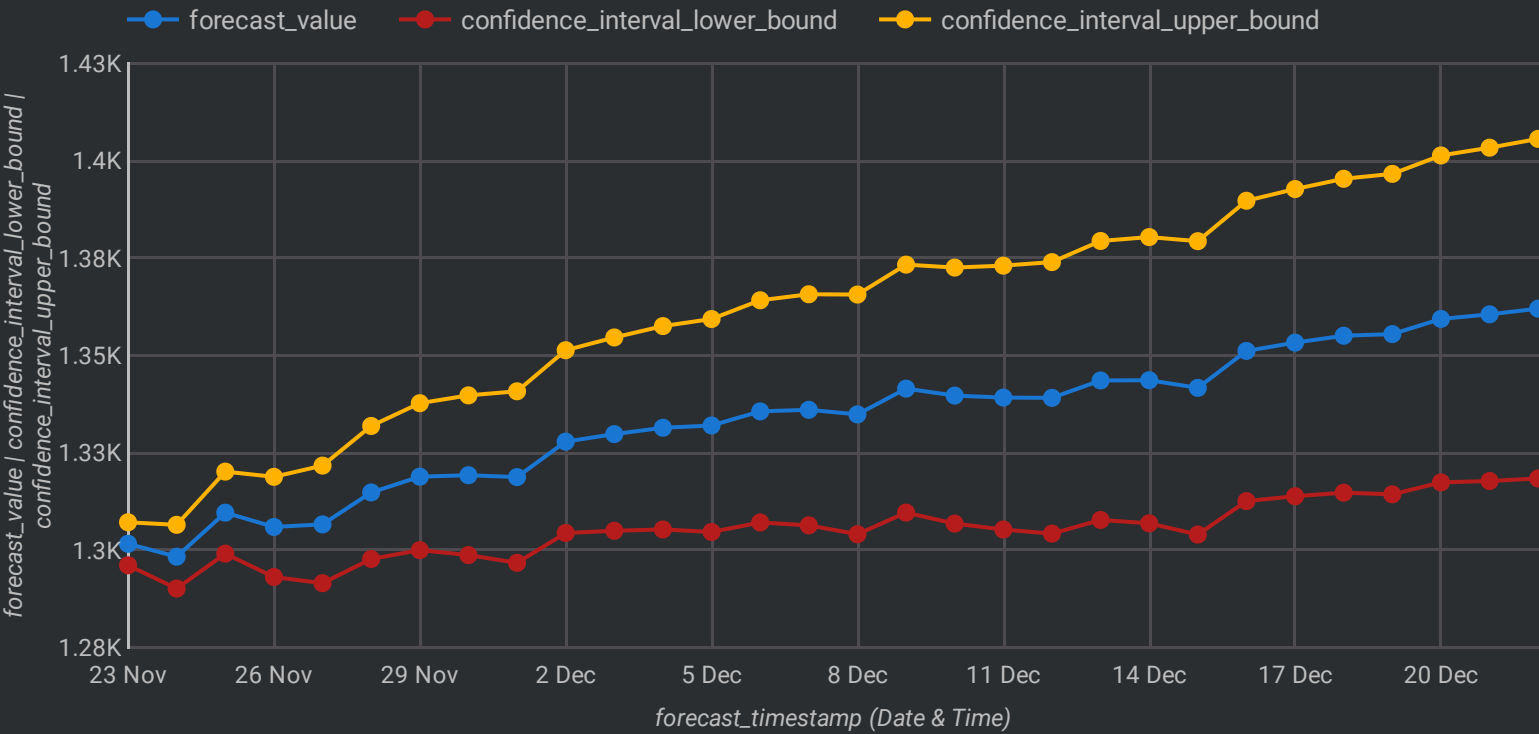
# Model Predictions

Trading Volume Predictions Using Random Forest



symbol
TCS
INFY
IBM
ACN

Price Predictions for the Next 30 days Using ARIMA

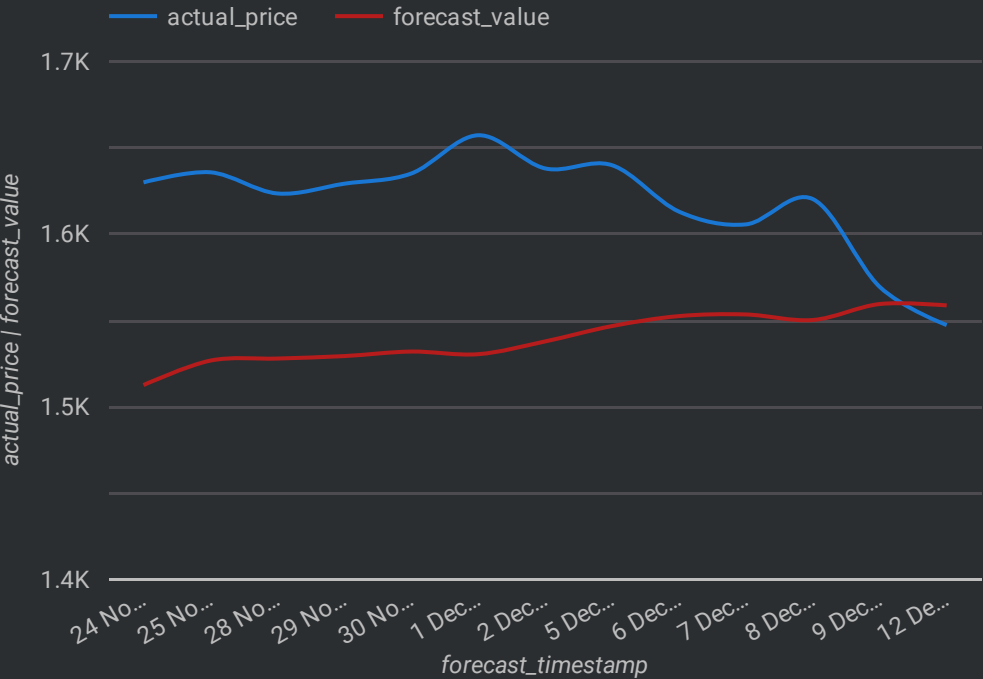




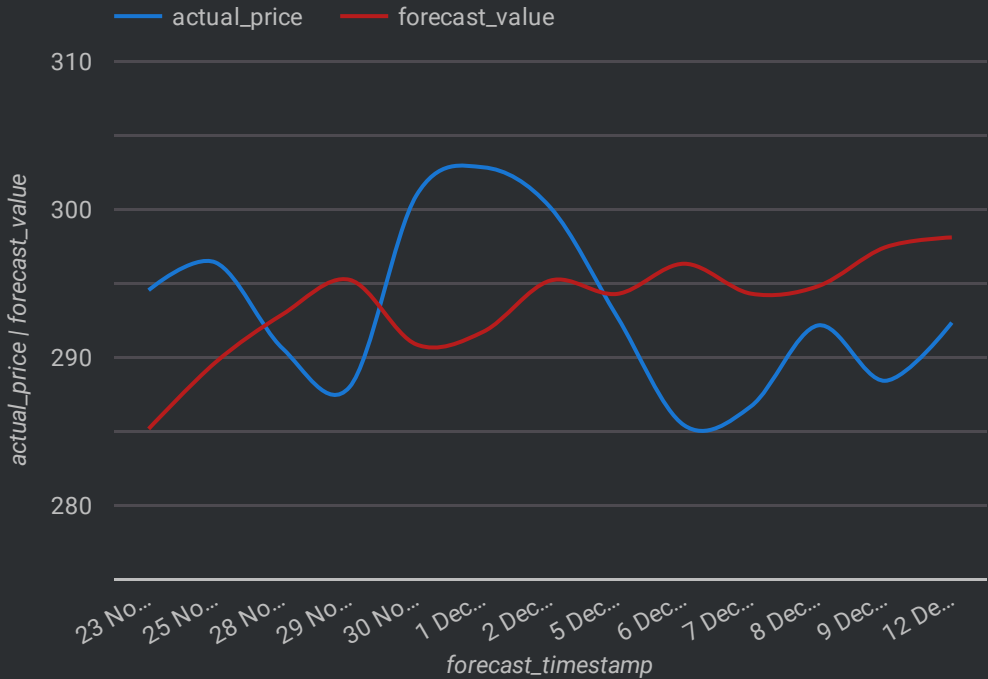
# Forecast vs Actuals



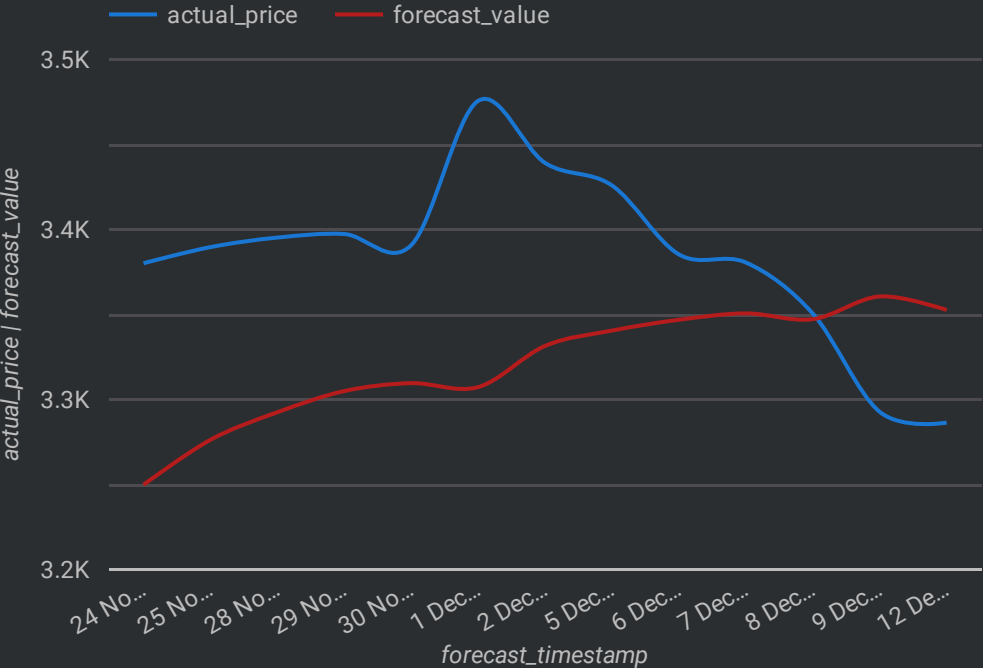
Infosys (INFY)



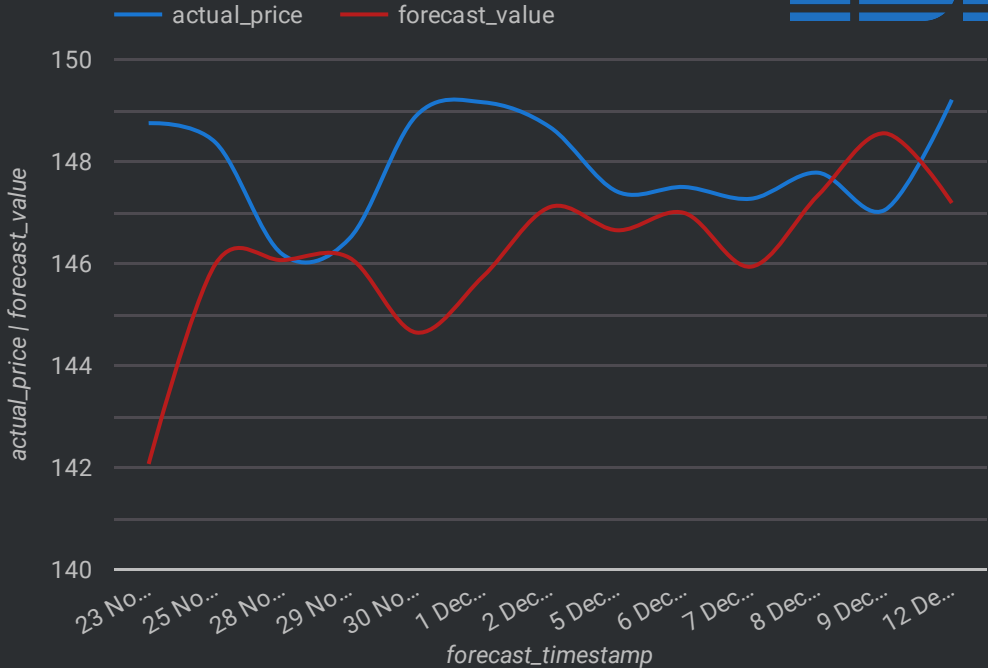
Accenture (ACN)  **accenture**



Tata Consultancy Services (TCS)



IBM (IBM)



# Price Variation - Risks & Opportunities

- Greater variation in stock prices means greater unpredictability and risk
- Prediction trends allow us to generalize the stock price variations
- Similarly, stocks with lesser variations would be more predictable and safer options
- **Infosys** and **TCS** are highly variable stocks. They are high-risk and less predictable
- **Accenture** and **IBM** are less variable stocks. They are low risk and more predictable

# Recommendations

For institutional and retail investors, it becomes essential that they know the trends for the stocks and make their trades accordingly.

- Our predictions help such entities formulate strategies corresponding to the stocks we have chosen for the short run.
- As per the models that we implemented, it would be prudent for an investor to use the predictions for hedging by making use of a protective put in the short run



# Assignment Feedback

The assignment was challenging and allowed us to work on all the aspects we covered in class. We were also excited to work on a new dataset. Yahoo Finance dataset was a good chance for us to learn more about the stock market.

The assignment did not take as long as we were well versed with the data pipeline creation. We only needed to spend time on creating our analysis. We spent nearly a week working on this project.

The skills we applied was GCS tools and material on financial investments.