

天時
地利
人和

Predicting the future

Using regression models to
forecast the daily closing price
of AUDUSD

Presented by Algotraders



天時

Heavenly timing,

地利

Ground advantage,

人和

Peace of mind



Our Team - Algotraders



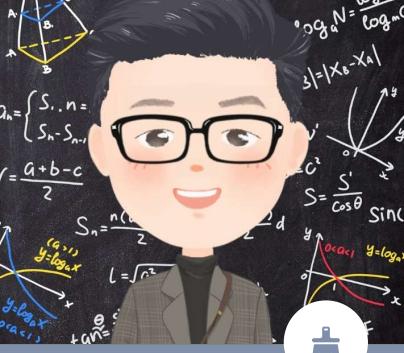
Simon

[Twitter](#) [Facebook](#) [Google+](#)



Venkat

[Twitter](#) [Facebook](#) [Google+](#)



Victor

[Twitter](#) [Facebook](#) [Google+](#)

Mathematical diagrams and formulas visible in the background:

$$a_n = \begin{cases} S_1, n=1 \\ S_n - S_{n-1}, n > 1 \end{cases}$$
$$r = \frac{a+b-c}{2}$$
$$S_n = \frac{n(a+b)}{2}$$
$$L = \sqrt{c^2}$$
$$y = \log_a N = \frac{\log_a b}{\log_a c}$$
$$S = \frac{S'}{\cos \theta} \sin \theta$$
$$y = \log_a x$$
$$\log_a N = \frac{\log_a b}{\log_a c}$$


Gary

[Twitter](#) [Facebook](#) [Google+](#)

Data Analytics Lifecycle

PLANNING ANALYTICS

What's your plan?

Understand the status of your business

PRESCRIPTIVE ANALYTICS

What should you do?

Finally, use prescriptive analytics to decide what you should do to overcome operational obstacles found in your analysis.



DESCRIPTIVE ANALYTICS

What happened?

Identify what happened and determine how you're performing against the plan.

DIAGNOSTIC ANALYTICS

Why did it happen?

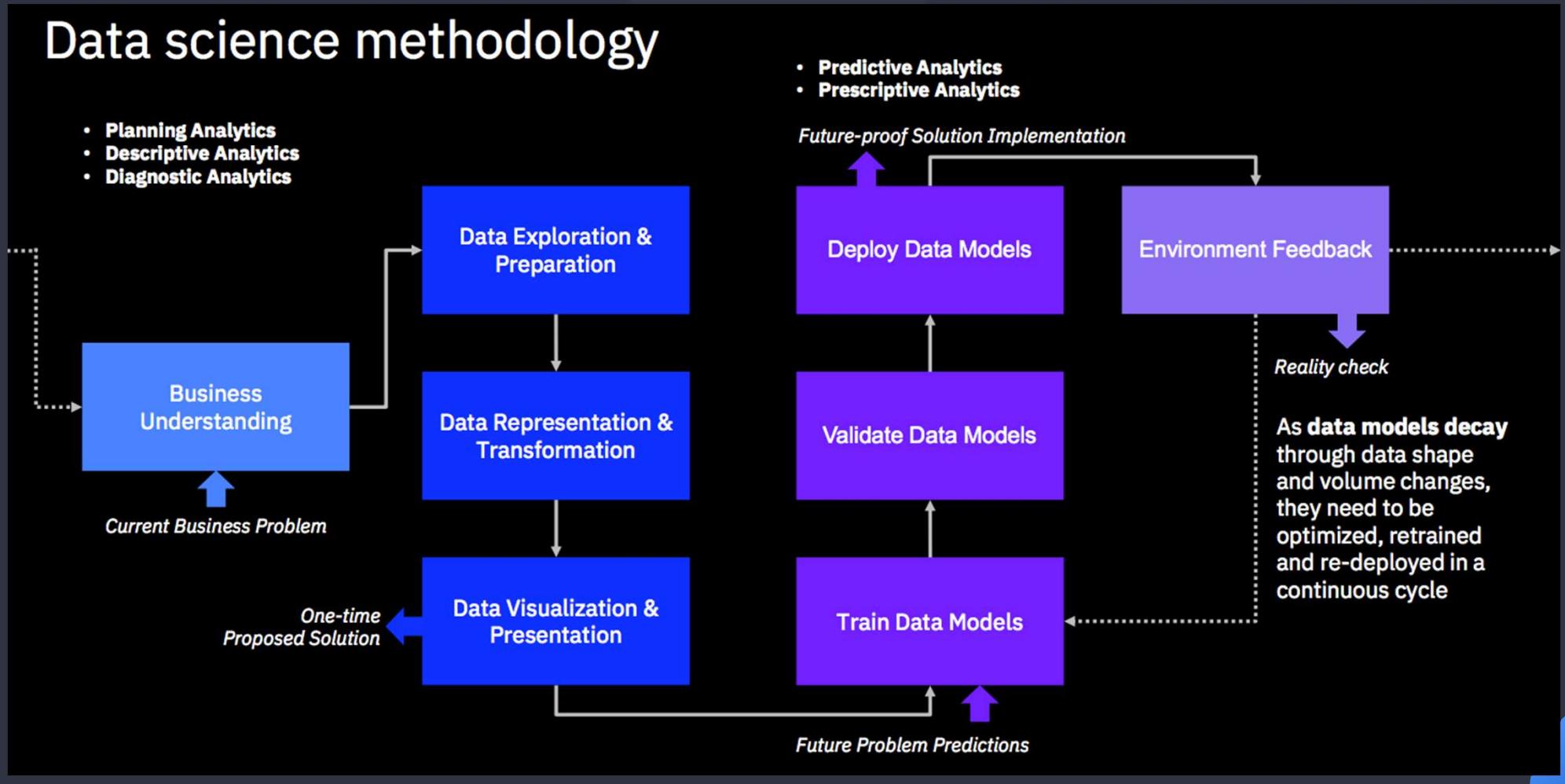
Then, diagnose why problems happened and identify key patterns

PREDICTIVE ANALYTICS

What happens next?

Use those patterns to predict future trends

8 Steps Data Science Methodology

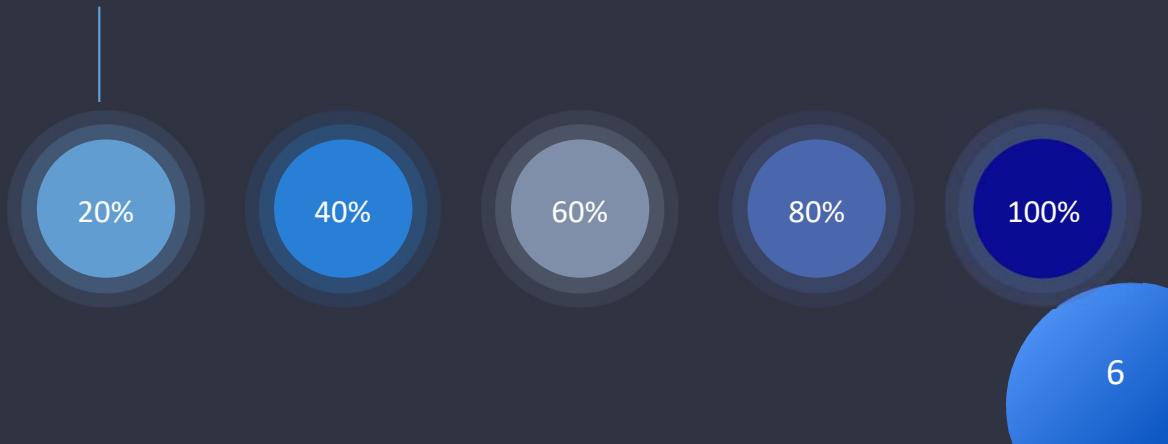


01

Business Understanding



Planning Analytics





GLOBAL FOREX MARKETS



\$6.6 trillion

daily trading volume



29%

increase since 2016



larger than
Futures market

larger than
Equities market

LARGEST FINANCIAL CENTRES



London



New York



Singapore

compareforexbrokers.com



MOST POPULAR FOREX PAIRS



USD/AUD

5.2%



USD/EUR

24%



USD/CAD

4.3%



USD/JPY

17.8%



USD/GBP

9.3%



USD/CNY

3.8%

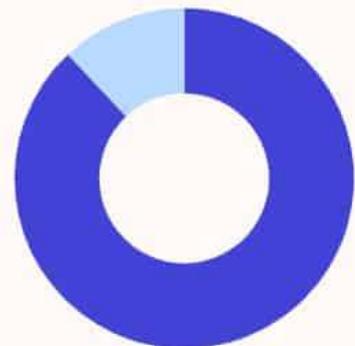


USD/CHF

3.6%

compareforexbrokers.com

FOREX MARKET TURNOVER BY CURRENCY



88% of forex trades involve the United States Dollar (USD)

EUR
32.3%

JPY
16.8%

GBP
12.8%

AUD
6.8%

CAD
5%

CHF
5%

compareforexbrokers.com

FOREX TRADER DEMOGRAPHICS



- Male Traders 89.1%
- Female Traders 10.9%



Aged between
25 - 34 years



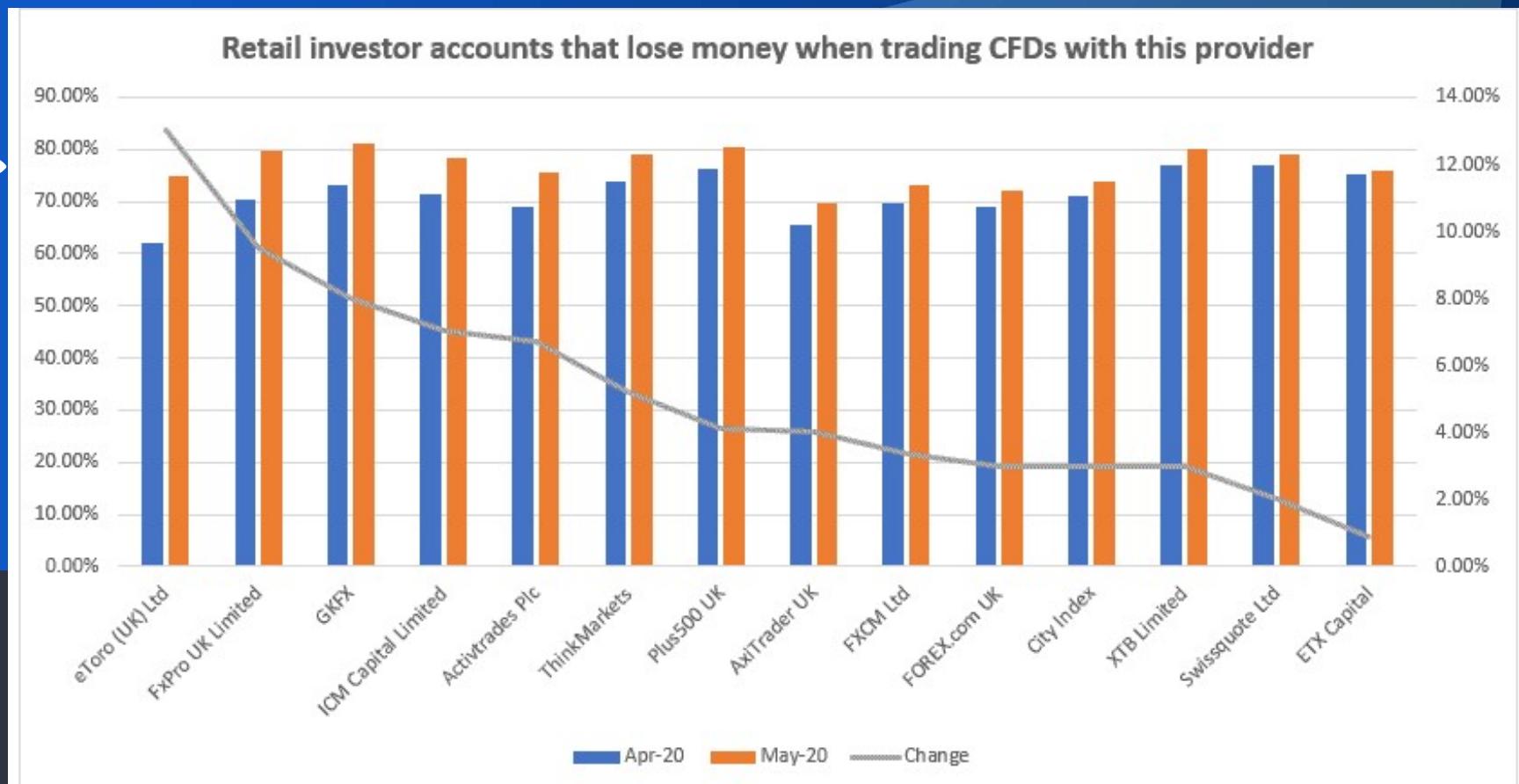
Aged between
34 - 45 years

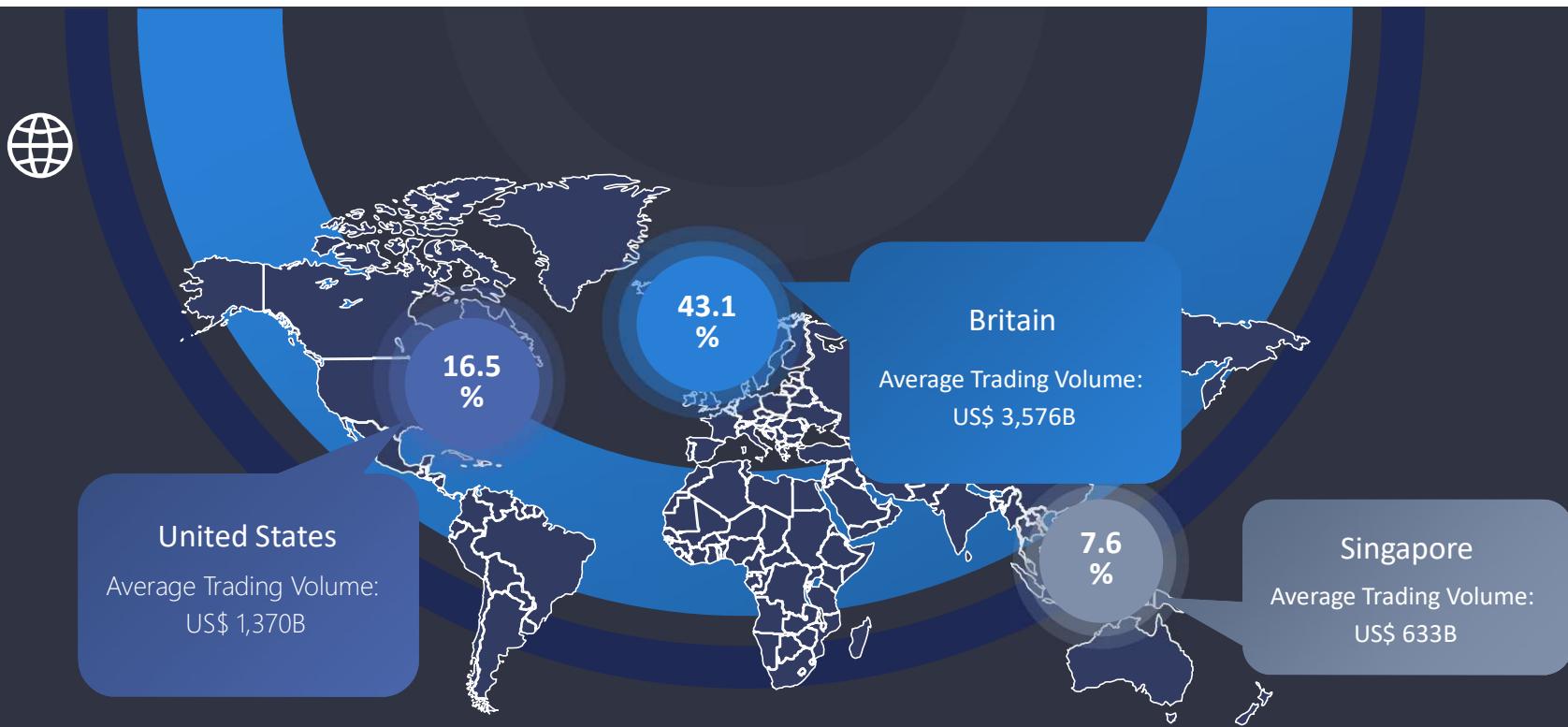


Aged 45 + years

compareforexbrokers.com

Retail Traders CFD Loss





TOP THREE

Global Forex Centre

“Singapore is third-largest forex centre globally and tops in Asia: Survey”
The Straits Times, 17 Sep 2019

Business Opportunity

We will improve the accuracy margin of individual day traders in Foreign Exchange.

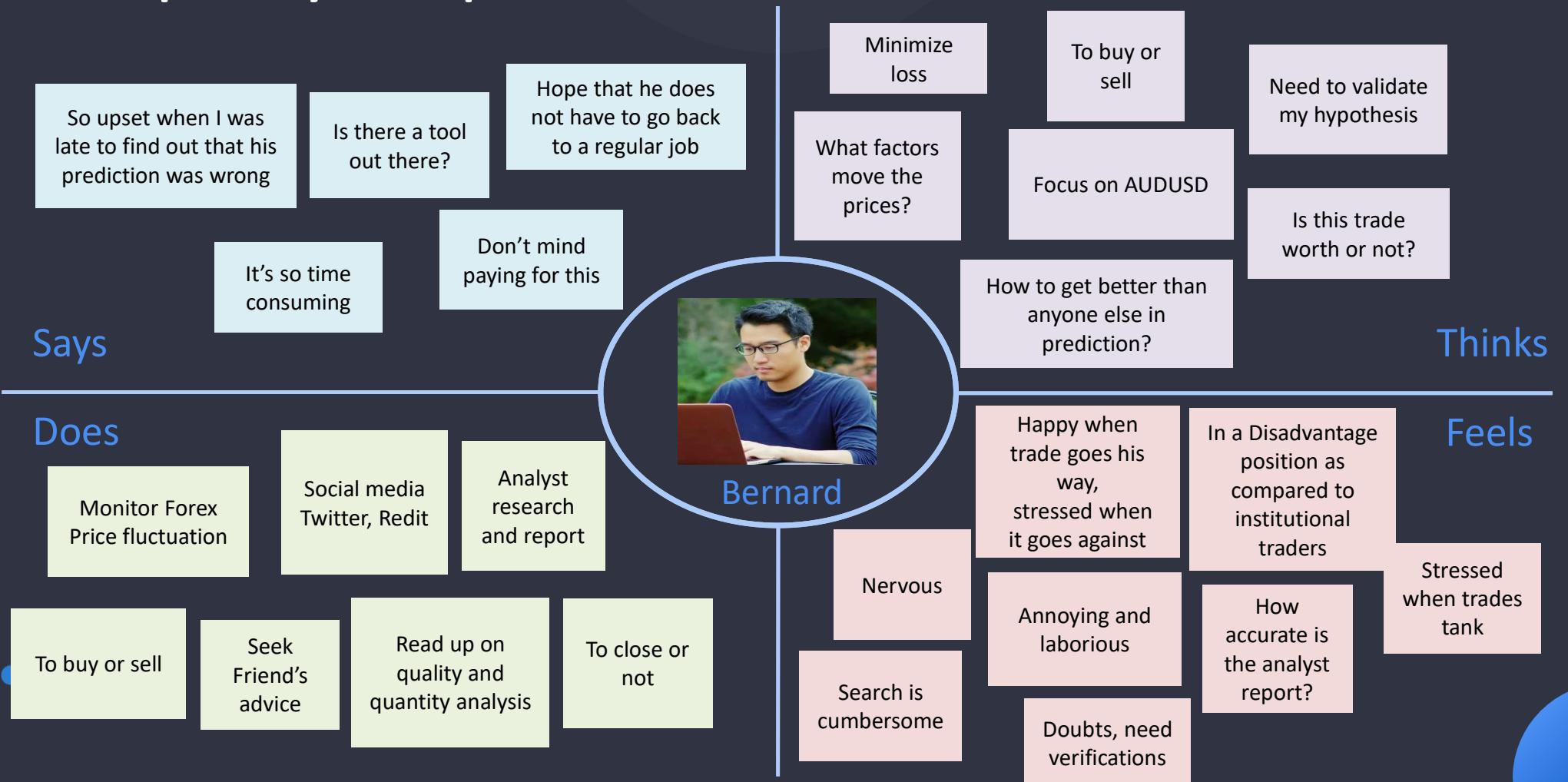
In the simplest term anyone unfamiliar with our industry would understand is, day traders need a better way to make prediction on movement of foreign currency pair

Our user is a man, 33 years old, a full-time day trader, who has done forex trading for 2 years.

This user struggles because he has low accuracy of predicting the currency movement using existing quantitative and qualitative analysis methods.

This would be great for our company because we could sell our data predictive services to individual forex day traders.

Empathy Map



Persona

Bernard

- Full-time Day Trader
- 33 years old
- University education in Australia
- Married, no kids
- Stays in a condominium
- Drives a car
- 2 years of Forex trading experience



Characteristics

Motivated, Intelligent

Concerns

How to make better prediction of the forex market

Wants

Increase his prediction accuracy
Make more money
Saves time
Early retirement

Hills

Who: Bernard, who is a Forex day trader

What: Will now have a better currency prediction tool on his hands to make forex trades

Wow: To provide him with better trading outcome statistically, helping him to make more money and achieve his goal of early retirement.



Hypothesis

Day trader will subscribe to an app for Forex prediction to be pushed to him automatically on every trading day.

Day trader understands the prediction is not a guarantee of making a profit, just like weather prediction is not fool-proof.



MVP Statement

If we provide: Day traders

With: Our solution that will push daily forex predictions to user, customised to each individual's requirements, giving the user advantage of better trading outcome

We will address the risk of: Accurately predicting the direction of currency movement

By measuring: Percentage of accurate predictions

We'll know we've arrived when we observe: There is over 65% accuracy in the predictions.

Goals

To offer user a better accuracy in forex prediction

To offer app as a service for financial institutions

Future Goals

To offer the services to include more currency pairs

To offer analytics tools and monitoring on social media

02

Data Exploration & Preparation

Descriptive Analytics

20%

40%

60%

80%

100%

20

Project Overview

The objective is to predict the exchange rate of AUDUSD based on:

- Australia Government Bond, 2 Years
- US Government Bond, 2 Years
- ASX200 Index
- S&P 500 Index
- USDX
- Interest Rate Differential

TRAINING DATASET

	Date	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
0	2011-01-02	1.02120	0.61	5.100	4742.5	127.05	79.13	4.490
1	2011-01-03	1.01665	0.61	5.100	4742.5	127.05	79.13	4.490
2	2011-01-04	1.00691	0.63	5.100	4742.5	126.98	79.44	4.470
3	2011-01-05	0.99974	0.71	5.030	4714.9	127.64	80.26	4.320
4	2011-01-06	0.99287	0.68	5.040	4725.0	127.39	80.79	4.360
...
3123	2020-12-27	0.76100	0.13	0.080	6664.8	369.00	90.32	-0.050
3124	2020-12-28	0.75871	0.13	0.080	6664.8	372.17	90.34	-0.050
3125	2020-12-29	0.76097	0.12	0.075	6700.3	371.46	89.99	-0.045
3126	2020-12-30	0.76874	0.12	0.080	6682.4	371.99	89.68	-0.040
3127	2020-12-31	0.76928	0.13	0.075	6587.1	373.88	89.94	-0.055

3128 rows × 8 columns

Drivers of AUDUSD

- The Reserve Bank of Australia states that interest rate differential as a main driver of Australian dollar

Source:
<https://www.rba.gov.au/education/resources/explainers/drivers-of-the-aud-exchange-rate.html>

In Education

Resources

Collections

In a Nutshell

Explainers

Digital Interactives

Videos

Learning Activities

Presentations

Snapshots

Chart Pack and Data

Posters

Teacher Updates

Delivery Notes

Illustrators

Book a Talk

Special Events for Students

Professional Development

Explore Economics

About the Program

Contact Us

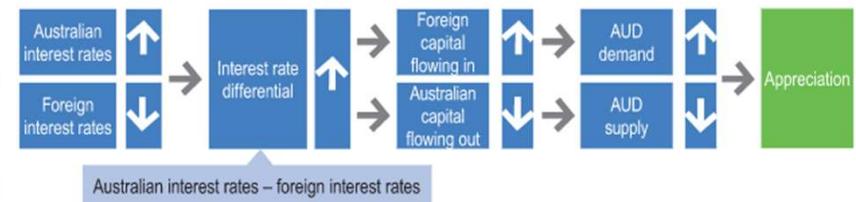
Education Program Updates

Drivers of the Australian Dollar Exchange Rate

Download the complete [Explainer PDF](#) 378KB

Australia has a floating exchange rate, which means that movements in the Australian dollar exchange rate are determined by the demand for, and supply of, Australian dollars in the foreign exchange market. There are a number of factors that affect demand and supply in this market. Some factors have longer-term effects on the value of the Australian dollar, while others influence its value over shorter periods of time. This explainer highlights some of the key drivers of the Australian dollar exchange rate. It also discusses foreign exchange intervention and the circumstances in which the Reserve Bank of Australia (RBA) might decide to intervene in the foreign exchange market.

Effect of an increase in Australia's interest rate differential



Longer-term Drivers

Interest rate differentials and capital flows

Australia's interest rate differential measures the difference between interest rates in Australia and those in other economies. The interest rate differential is a key driver of the demand for, and supply of, Australian dollars. It is also an important driver of capital flows, which measure the money that flows into, and out of, Australia for investment purposes. (See [Explainer: The Balance of Payments](#).)

For the Australian dollar, the focus is typically on the difference between Australian interest rates and those in the major advanced economies, such as the United States (US), Europe and Japan. Because the interest rate differential is a key driver of exchange rates, the RBA's monetary policy decisions play a key role in influencing the exchange rate. (See [Explainer: The Transmission of Monetary Policy](#) and [Explainer: Bonds and the Yield Curve](#) for a discussion of how monetary policy affects interest rates and the exchange rate.)

Data Dictionary

1. AUDUSD is the abbreviation for the Australian dollar and U.S. dollar currency pair or cross. A currency pair tells the reader how much of one currency is needed to purchase one unit of another currency.
2. US Treasury Yield Curve Rate (eg. US Bond 2 Yr) is the spot rate Treasury curve is a yield curve constructed using Treasury spot rates rather than yields.
3. Australian Govt, Bonds (eg. AUS Bond 2 Yr) are a type of instrument for investment that are the safest type of bonds. If you buy and hold them to maturity (2, 3, 5, 10 years), you're guaranteed a rate of return.
4. ASX 200 Index is the benchmark institutional investable stock market index in Australia, comprising the 200 largest stocks by float-adjusted market capitalization.
5. S&P 500 (SPY)- SPDR is an acronym for the Standard & Poor's Depository Receipts. It is an Exchange Traded Fund. It is designed to track the S&P 500 stock market index. This fund is the largest ETF in the world.
6. US Dollar Index (USDX) - The US Dollar Index is used to measure the value of the dollar against a basket of six world currencies - Euro, Swiss Franc, Japanese Yen, Canadian dollar, British pound, and Swedish Krona.
7. Interest Rate Differential (IRD) is a difference in the interest rate between two government bonds. If one has an interest rate of 3% and the other has an interest rate of 1%, it has a 2% interest rate differential.

Data Sources

1. AUDUSD, daily: Forex Academy, <https://eatradingacademy.com/software/forex-historical-data/>
2. US Treasury Yield Curve Rate, daily: US Department of Treasury, <https://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/TextView.aspx?data=yieldAll>
3. Australian Govt, Bonds, daily: The Reserve Bank of Australia, <https://www.rba.gov.au/statistics/historical-data.html>
4. ASX 200 Index, daily: Investing.com, <https://www.investing.com/indices/aus-200-historical-data>.
5. S&P 500 (SPY), daily: Investing.com, <https://www.investing.com/etfs/spdr-s-p-500-historical-data>.
6. US Dollar Index, daily: Investing.com, <https://www.investing.com/indices/usdollar-historical-data>.
7. Interest Rate Differential (IRD) **Assumption:** Differences in AUS and US government bonds interest rates will have an impact on the AUDUSD currency pair. Calculation based on:

Interest Rate Differential (IRD) = AUS Bond 2 Yr, daily – US Treasury Bond 2Yr, daily

Data Cleansing & Preparation

- Convert original string to column (AUDUSD)
- Delete not useful columns: ‘Open’, ‘High’, ‘Low’, ‘Volume’ and ‘Time Stamp’
- Standardise ‘Date’ format
- Create master Merged dataset
- Y2011 to Y2020 (Training Data)
- Y2021 (Holdout Data)
- To Remove or Impute Missing Values
- Calculating IRD

Tool: Microsoft Excel with Power Query add-on

The screenshot shows a Microsoft Excel spreadsheet titled "AUDUSD1440.csv". The top ribbon is visible with tabs like Home, Insert, Page Layout, Formulas, Data, Review, etc. A red circle highlights the "Data" tab. Below it, the "Text to Columns" wizard is open, Step 1 of 3, with the "Delimited" option selected. The preview pane shows the first few rows of data. Step 2 of 3 is also visible, showing the "Text qualifier" dropdown set to "None". Step 3 of 3 shows the "Advanced" options for data types, with "Text" selected for the date column. The bottom part of the screen shows the main Excel worksheet with data from April 2011 to March 2021. A red box highlights the "Delete" option in the context menu for a selected cell.

Master Data

- AUDUSD: 3,128 rows
- US Bond 2Yr: 2,501 rows
- AUS Bond 2Yr: 2,525 rows
- ASX 200: 2,579 rows
- S&P 500: 2,517 rows
- USDX: 2,610 rows
- Dates are different due to holidays & other differences

Screenshot of a Microsoft Excel spreadsheet titled "AUDUSD1446-2011-2020_combined (1).xlsx". The spreadsheet contains data from various financial markets across different years. Red arrows highlight specific data points in the "Date" column.

Key columns visible:

- Date
- AUDUSD Close
- US Bond 2Yr
- AUS Bond 2Yr
- ASX200
- SP500
- USDX

Red Arrows:

- An arrow points from row 2502 (Date: 30/12/18) to row 31/12/20 (Date: 31/12/20).
- An arrow points from row 2518 (Date: 27/12/19) to row 31/12/20 (Date: 31/12/20).
- An arrow points from row 2526 (Date: 27/12/19) to row 31/12/20 (Date: 31/12/20).
- An arrow points from row 2580 (Date: 30/12/19) to row 31/12/20 (Date: 31/12/20).
- An arrow points from row 3125 (Date: 27/12/20) to row 31/12/20 (Date: 31/12/20).

Training Data (Remove)

- Remove missing values for missing dates
 - Final training dataset: 2,443 rows, 8 columns

Training Data (Impute)

- Impute missing values for missing dates
- Final training dataset: 3,128 rows, 8 columns

A	B	C	D	E	F	G	H
1	Date	AUDUSD Close	US Bond 2Yr	Aus Bond 2 Yr	ASX 200	SP 500	USDX
2	2/1/2011	1.0212	0.61	5.1	4742.5	127.05	79.13
3	3/1/2011	1.01665	0.61	5.1	4742.5	127.05	79.13
4	4/1/2011	1.00691	0.63	5.1	4742.5	126.98	79.44
5	5/1/2011	0.99974	0.71	5.03	4714.9	127.64	80.26
6	6/1/2011	0.99287	0.68	5.04	4725	127.39	80.79
7	7/1/2011	0.99558	0.6	5.03	4705	127.14	81.01
8	9/1/2011	0.99512	0.6	5.03	4705	127.14	81.01
9	10/1/2011	0.99465	0.59	4.985	4712.3	126.98	80.88
10	11/1/2011	0.98218	0.6	4.84	4710.7	127.43	80.85
11	12/1/2011	0.99571	0.61	4.89	4724.2	128.58	80.03
12	13/1/2011	0.99791	0.59	4.995	4795.2	128.37	79.19
13	14/1/2011	0.98844	0.59	4.98	4801.5	129.3	79.16
14	16/1/2011	0.99069	0.59	4.98	4801.5	129.3	79.16
15	17/1/2011	0.99272	0.59	4.985	4763.1	129.3	79.34
16	18/1/2011	0.99738	0.6	4.97	4801.8	129.52	78.96
17	19/1/2011	0.99678	0.6	5.035	4834.6	128.25	78.64
18	20/1/2011	0.98543	0.65	4.985	4783.7	128.08	78.82
19	21/1/2011	0.98962	0.63	5.04	4755.7	128.37	78.21
20	23/1/2011	0.98937	0.63	5.04	4755.7	128.37	78.21
21	24/1/2011	0.99743	0.65	5.04	4786	129.1	78.05
22	25/1/2011	0.99504	0.62	4.96	4807.8	129.17	78
23	26/1/2011	0.99936	0.62	4.96	4807.8	129.67	77.9
24	27/1/2011	0.98955	0.59	4.925	4806.1	129.99	77.73
25	28/1/2011	0.99381	0.54	4.88	4774.9	127.72	78.13
26	30/1/2011	0.98685	0.54	4.88	4774.9	127.72	78.13
27	31/1/2011	0.99711	0.58	4.885	4753.9	128.68	77.73
28	1/2/2011	1.01103	0.61	4.91	4752.1	130.74	77.07
29	2/2/2011	1.01136	0.67	4.915	4796.5	130.49	77.16
30	3/2/2011	1.01453	0.71	4.975	4820.6	130.78	77.75
31	4/2/2011	1.01368	0.77	5.095	4862.7	131.15	78.04
3111	10/12/2020	0.75228	0.14	0.085	6683.1	366.73	90.82
3112	11/12/2020	0.75305	0.11	0.09	6642.6	366.3	90.98
3113	13/12/2020	0.75494	0.11	0.09	6642.6	366.3	90.98
3114	14/12/2020	0.75346	0.13	0.105	6660.2	364.66	90.71
3115	15/12/2020	0.75572	0.11	0.095	6631.3	369.59	90.47
3116	16/12/2020	0.75692	0.13	0.1	6679.2	370.17	90.45
3117	17/12/2020	0.76139	0.13	0.1	6756.7	372.24	89.82
3118	18/12/2020	0.76193	0.13	0.105	6675.5	369.18	90.02
3119	20/12/2020	0.75868	0.13	0.105	6675.5	369.18	90.02
3120	21/12/2020	0.75753	0.13	0.09	6669.9	367.86	90.04
3121	22/12/2020	0.75536	0.13	0.085	6599.6	367.24	90.65
3122	23/12/2020	0.75822	0.13	0.085	6643.1	367.57	90.41
3123	24/12/2020	0.75919	0.13	0.08	6664.8	369	90.32
3124	25/12/2020	0.7604	0.13	0.08	6664.8	369	90.32
3125	27/12/2020	0.761	0.13	0.08	6664.8	369	90.32
3126	28/12/2020	0.75871	0.13	0.08	6664.8	372.17	90.34
3127	29/12/2020	0.76097	0.12	0.075	6700.3	371.46	89.99
3128	30/12/2020	0.76874	0.12	0.08	6682.4	371.99	89.68
3129	31/12/2020	0.76928	0.13	0.075	6587.1	373.88	89.94
3130							
3131							
3132							
3133							
3134							
3135							
3136							



Holdout Data (Remove)

- Remove missing values for missing dates
- Final holdout dataset is: 43 rows, 8 columns

AUDUSD_2021 Combined - Excel

TABLE TOOLS

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW POWER QUERY QUERY DESIGN

From Access From Web From Other Sources Existing Connections Refresh All Properties Connections Sort Filter Advanced Sort & Filter Text to Columns Flash Fill Remove Duplicates Data Tools Consolidate What-If Analysis

H3 =[@[AUS Bond 2Yr]]-[@[US Bond 2Yr]]

	Date	AUDUSD Close	US Bond 2Yr	AUS Bond 2Yr	ASX200	SP500	USDX	IRD	
1	1/4/2021	0.76692	0.17	0.08	6684.2	368.79	89.87	-0.09	
2	1/13/2021	0.77461	0.14	0.08	6686.6	379.79	90.36	-0.06	
3	1/14/2021	0.77752	0.16	0.08	6715.3	378.46	90.24	-0.08	
4	1/15/2021	0.7698	0.13	0.075	6715.4	375.7	90.77	-0.055	
5	1/19/2021	0.77131	0.14	0.075	6742.6	378.65	90.5	-0.065	
6	1/20/2021	0.77544	0.13	0.075	6770.4	383.89	90.47	-0.055	
7	1/21/2021	0.77593	0.13	0.075	6823.7	384.24	90.13	-0.055	
8	1/22/2021	0.77117	0.13	0.08	6800.4	382.88	90.24	-0.05	
9	1/25/2021	0.77143	0.13	0.08	6824.7	384.39	90.39	-0.05	
10	1/27/2021	0.76478	0.12	0.085	6780.6	374.41	90.65	-0.035	
11	1/28/2021	0.76703	0.12	0.085	6649.7	377.63	90.46	-0.035	
12	1/29/2021	0.76405	0.11	0.085	6607.4	370.07	90.58	-0.025	
13	2/2/2021	0.76055	0.11	0.09	6762.6	381.55	91.2	-0.02	
14	2/3/2021	0.76359	0.13	0.09	6824.6	381.85	91.17	-0.04	
15	2/4/2021	0.75999	0.19	0.095	6765.5	386.19	91.53	-0.095	
16	2/16/2021	0.7735	0.13	0.09	6917.3	392.3	90.51	-0.04	
17	2/17/2021	0.77576	0.11	0.095	6885.2	392.39	90.95	-0.015	
18	2/18/2021	0.77678	0.11	0.105	6885.9	390.72	90.59	-0.005	
19	2/19/2021	0.78658	0.11	0.115	6793.8	390.03	90.36	0.005	
20	2/22/2021	0.79137	0.11	0.12	6780.9	387.03	90.01	0.01	
21	2/23/2021	0.79067	0.11	0.12	6839.2	387.5	90.17	0.01	
22	2/24/2021	0.79752	0.12	0.115	6777.8	391.77	90.18	-0.005	
23	2/25/2021	0.78622	0.17	0.11	6834	382.33	90.13	-0.06	
24	2/26/2021	0.77034	0.14	0.105	6673.3	380.36	90.88	-0.035	
25	3/2/2021	0.78238	0.11	0.1	6762.3	386.54	90.79	-0.01	
26	3/3/2021	0.77525	0.14	0.11	6818	381.42	90.95	-0.03	
27	3/15/2021	0.77475	0.14	0.085	6773	396.41	91.83	-0.055	
28	3/16/2021	0.77457	0.15	0.08	6827.1	395.91	91.86	-0.07	
29	3/17/2021	0.7814	0.13	0.08	6795.2	397.26	91.44	-0.05	
30	3/18/2021	0.7752	0.16	0.085	6745.9	391.48	91.86	-0.075	
31	3/19/2021	0.77369	0.16	0.085	6708.2	389.48	91.92	-0.075	
32	3/22/2021	0.77443	0.15	0.08	6752.5	392.59	91.74	-0.07	
33	3/23/2021	0.76062	0.15	0.08	6745.4	389.5	92.34	-0.07	
34	3/24/2021	0.75903	0.14	0.075	6778.8	387.52	92.53	-0.065	
35	3/25/2021	0.75921	0.14	0.07	6790.6	389.7	92.53	-0.07	
36	3/26/2021	0.76371	0.14	0.07	6824.2	395.98	92.77	-0.07	
37	3/29/2021	0.76381	0.14	0.075	6799.5	395.78	92.94	-0.065	
38	3/30/2021	0.76078	0.16	0.07	6738.4	394.73	93.3	-0.09	
39	3/31/2021	0.75935	0.16	0.075	6790.7	396.33	93.23	-0.085	
40	4/1/2021	0.76161	0.11	0.075	6828.7	400.61	92.93	-0.035	
41	4/13/2021	0.76467	0.16	0.07	6976.9	412.86	91.85	-0.09	
42	4/14/2021	0.77244	0.16	0.07	7023.1	411.45	91.69	-0.09	
43	4/15/2021	0.77148	0.16	0.075	7058.6	415.87	91.69	-0.085	
44									
45									
46									



Holdout Data (Impute)

- Impute missing values for missing dates
- Final holdout dataset is: 89 rows, 8 columns



Merged New Data_2021_Imput Values - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW POWER QUERY

Normal Page Break Preview Page Layout Custom Gridlines Headings

Zoom 100% Zoom to Selection Window New Arrange Freeze All Panes Hide Unhide

Workbook Views Show Zoom Window

J91	A	B	C	D	E	F	G	H	I	L
1	Date	AUDUSD Close	US Bond 2Yr	AUS Bond 2Yr	ASX200	SP500	USDX	IRD		
2	1/3/2021	0.77109	0.13	0.08	6684.2	368.79	89.87	-0.05		
3	1/4/2021	0.76692	0.17	0.08	6684.2	368.79	89.87	-0.09		
4	1/5/2021	0.77618	0.17	0.07	6681.9	371.33	89.44	-0.1		
5	1/6/2021	0.78152	0.17	0.07	6607.1	373.55	89.53	-0.1		
6	1/7/2021	0.77682	0.17	0.075	6712	379.1	89.83	-0.095		
7	1/8/2021	0.77591	0.17	0.075	6757.9	381.26	90.1	-0.095		
8	1/10/2021	0.77202	0.17	0.075	6757.9	381.26	90.1	-0.095		
9	1/11/2021	0.77068	0.17	0.075	6697.2	378.69	90.46	-0.095		
10	1/12/2021	0.77752	0.17	0.075	6679.1	378.77	90.09	-0.095		
11	1/13/2021	0.77461	0.14	0.08	6686.6	379.79	90.36	-0.06		
12	1/14/2021	0.77752	0.16	0.08	6715.3	378.46	90.24	-0.08		
13	1/15/2021	0.7698	0.13	0.075	6715.4	375.7	90.77	-0.055		
14	1/17/2021	0.76828	0.13	0.075	6715.4	375.7	90.77	-0.055		
78	4/1/2021	0.76161	0.11	0.075	6828.7	400.61	92.93	-0.035		
79	4/2/2021	0.75976	0.11	0.075	6828.7	400.61	93.02	-0.035		
80	4/4/2021	0.76189	0.11	0.075	6828.7	400.61	93.02	-0.035		
81	4/5/2021	0.7657	0.11	0.075	6828.7	406.36	92.6	-0.035		
82	4/6/2021	0.76659	0.11	0.075	6885.9	406.12	92.33	-0.035		
83	4/7/2021	0.76131	0.11	0.075	6928	406.59	92.46	-0.035		
84	4/8/2021	0.76561	0.11	0.075	6998.8	408.52	92.06	-0.035		
85	4/9/2021	0.76168	0.11	0.07	6995.2	411.49	92.16	-0.04		
86	4/11/2021	0.7621	0.11	0.07	6995.2	411.49	92.16	-0.04		
87	4/12/2021	0.76265	0.11	0.07	6974	411.64	92.14	-0.04		
88	4/13/2021	0.76467	0.16	0.07	6976.9	412.86	91.85	-0.09		
89	4/14/2021	0.77244	0.16	0.07	7023.1	411.45	91.69	-0.09		
90	4/15/2021	0.77148	0.16	0.075	7058.6	415.87	91.69	-0.085		
91										

Merge5

READY

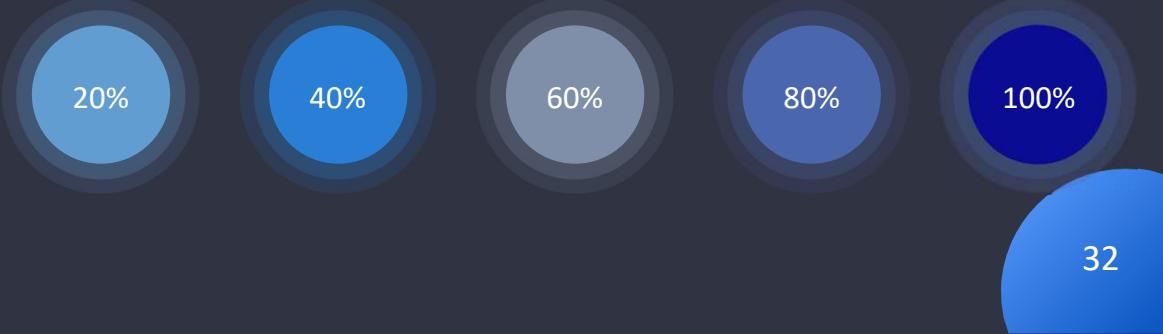
Details of Data Cleansing & Preparation Steps

Please refer to [Appendix Section](#)

03

Data Representation & Transformation

Descriptive Analytics



EXPLORATORY DATA ANALYSIS

EDA is a method to analyse data with the help of various tools and graphical techniques. After preparing and cleansing the data, we need to get to know the data by doing some Exploratory Analysis before making further analysis and predictions.

So descriptive statistics such as histograms helps to summarize and describe data in a meaningful way which allows for simple visual interpretation and understanding of data.

A better-organized and transformed data may be easier for both humans and computers to use. It can increase the efficiency of analytic and business processes and enable better data-driven decision-making.

For our EDA, we used Jupyter Notebook along with the necessary libraries to conduct the following Exploratory Data Analysis.

INITIAL INSPECTION

Perform EDA using Pandas Profiling

```
In [1]: import numpy as np  
import pandas as pd  
from pandas_profiling import ProfileReport
```

```
In [2]: df = pd.read_excel('Merged Data_Edited (1).xlsx')
```

```
In [3]: df.head()
```

```
out[3]:
```

	Date	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
0	2011-01-02	1.02120	0.61	5.10	4742.5	127.05	79.13	4.49
1	2011-01-03	1.01665	0.61	5.10	4742.5	127.05	79.13	4.49
2	2011-01-04	1.00691	0.63	5.10	4742.5	126.98	79.44	4.47
3	2011-01-05	0.99974	0.71	5.03	4714.9	127.64	80.26	4.32
4	2011-01-06	0.99287	0.68	5.04	4725.0	127.39	80.79	4.36

```
In [4]: df.tail()
```

```
Out[4]:
```

	Date	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
3123	2020-12-27	0.76100	0.13	0.080	6664.8	369.00	90.32	-0.050
3124	2020-12-28	0.75871	0.13	0.080	6664.8	372.17	90.34	-0.050
3125	2020-12-29	0.76097	0.12	0.075	6700.3	371.46	89.99	-0.045
3126	2020-12-30	0.76874	0.12	0.080	6682.4	371.99	89.68	-0.040
3127	2020-12-31	0.76928	0.13	0.075	6587.1	373.88	89.94	-0.055

- Importing of our Data Set
- Calling the first 5 records and last 5 records

Column Check

```
In [5]: profile = ProfileReport(df, title='Pandas Profiling Report', explorative=True)
```

```
In [6]: df.columns
```

```
Out[6]: Index(['Date', 'AUDUSD_Close', 'US_bond_2Yr', 'Aus_bond_2Yr', 'ASX200',
   'SP500', 'USDX', 'IRD'],
   dtype='object')
```

Checking the column header to ensure it's displaying properly

Dataframe Info

```
df.info()
```

#	Column	Non-Null Count	Dtype
0	Date	3128 non-null	datetime64[ns]
1	AUDUSD_Close	3128 non-null	float64
2	US_bond_2Yr	3128 non-null	float64
3	Aus_bond_2Yr	3128 non-null	float64
4	ASX200	3128 non-null	float64
5	SP500	3128 non-null	float64
6	USDX	3128 non-null	float64
7	IRD	3128 non-null	float64

dtypes: datetime64[ns](1), float64(7)
memory usage: 195.6 KB

The data frame info displays the no. Of rows and columns as well as the data types and if there is any missing values

OVERVIEW

Dataset statistics		Variable types	
Number of variables	8	Date Time	1
Number of observations	3128	Numeric	7
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	195.6 KiB		
Average record size in memory	64.0 B		

Running the pandas profiling on the Data Frame returns a profiling report. The overview section provides overall data set information and has 2 sub-sections – Dataset information and Variables Types.

04

Data Visualization & Presentation



Diagnostic Analytics

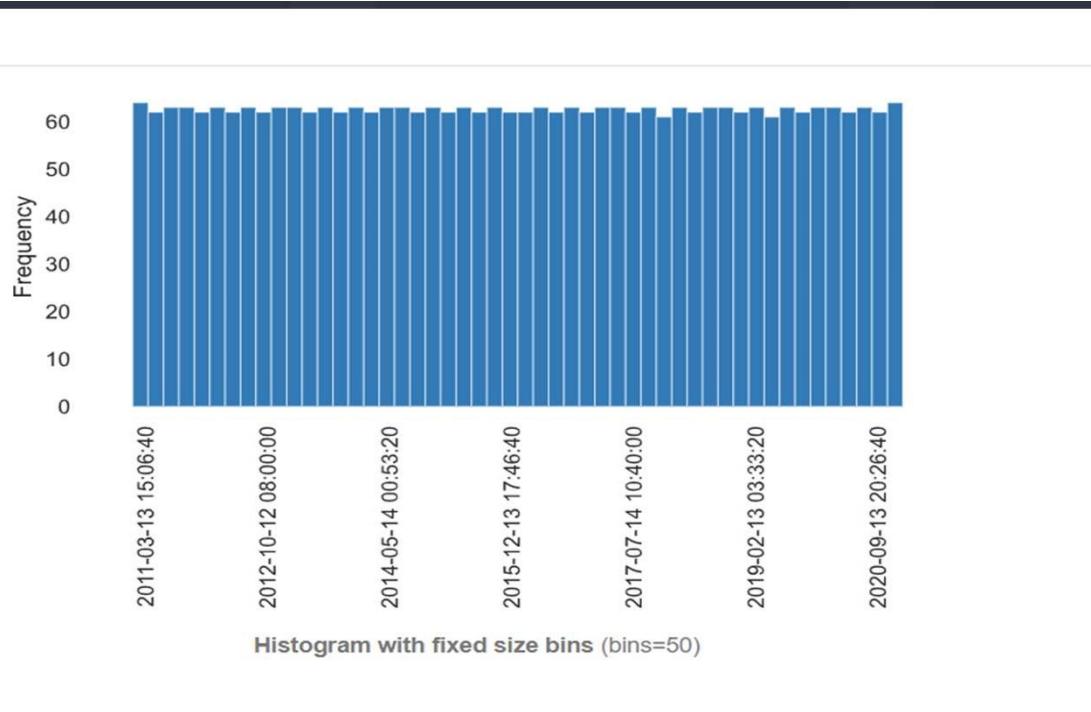


Data Visualization

Using Pandas Profiling, an open source python module, we can quickly do a pretty amazing EDA with just a few lines of code. It helps us to visualize and understand the distribution of each variable. Another amazing thing is that it also generates and outputs a report in html.

DATE HISTOGRAM AND STATISTICS

Shows the 10 year Data
From 02/01.2011 -
31/12/2020



Distinct	3128	Minimum	2011-01-02 00:00:00
Distinct (%)	100.0%	Maximum	2020-12-31 00:00:00
Missing	0		
Missing (%)	0.0%		
Memory size	24.6 KiB		

Date

Date

UNIQUE

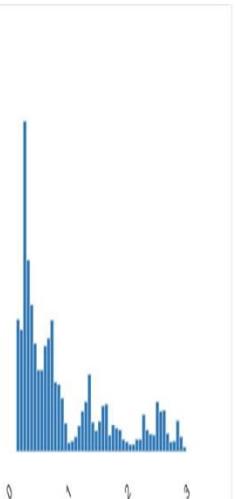
AUD/USD STATISTICS AND HISTOGRAM



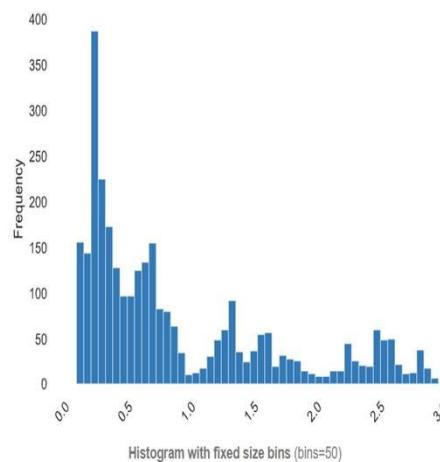
US BOND 2 YEARS STATISTICS AND HISTOGRAM

US_bond_2Yr
Real number (32)

Distinct	276	Mean	0.9297090793
Distinct (%)	8.8%	Minimum	0.11
Missing	0	Maximum	2.98
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	24.6 kB



Statistics Histogram Common values Extreme values



Statistics

Histogram

Common values

Extreme values

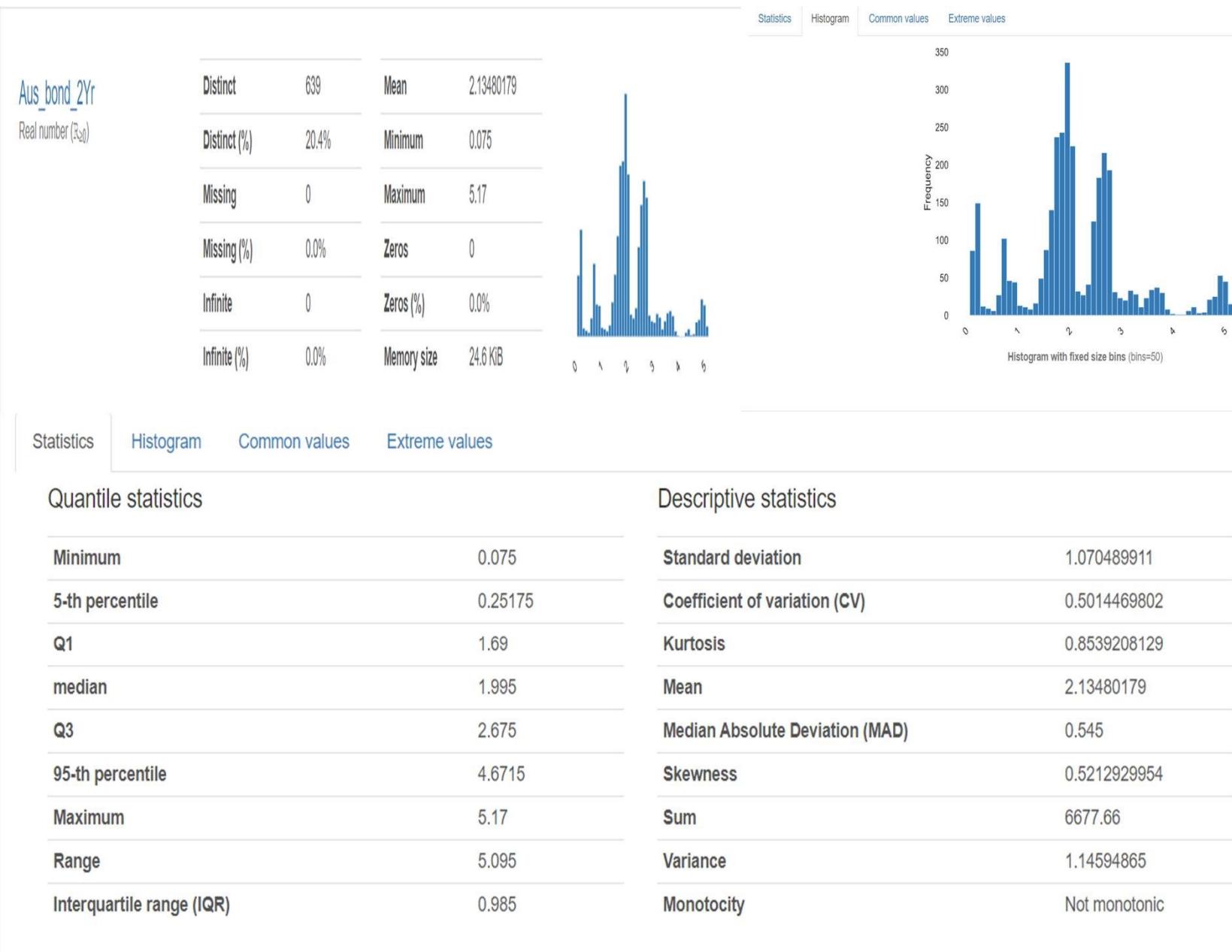
Quantile statistics

Minimum	0.11
5-th percentile	0.17
Q1	0.3
median	0.64
Q3	1.36
95-th percentile	2.58
Maximum	2.98
Range	2.87
Interquartile range (IQR)	1.06

Descriptive statistics

Standard deviation	0.7776076724
Coefficient of variation (CV)	0.8363989227
Kurtosis	-0.07796194461
Mean	0.9297090793
Median Absolute Deviation (MAD)	0.37
Skewness	1.060929246
Sum	2908.13
Variance	0.6046736921
Monotocity	Not monotonic

AUS BOND 2 YEARS STATISTICS AND HISTOGRAM



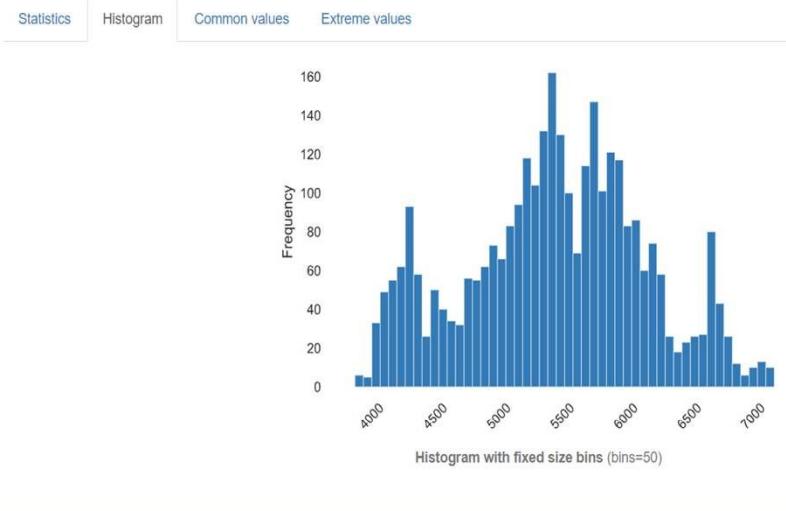
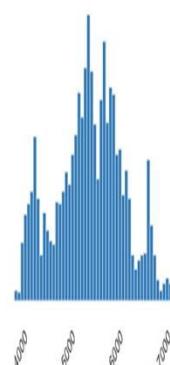
ASX 200 STATISTICS AND HISTOGRAM

ASX200

Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION

Distinct	249	Mean	5439.33141
Distinct (%)	78.3%	Minimum	3863.9
Missing	0	Maximum	7162.5
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	24.6 KB



Statistics

Histogram

Common values

Extreme values

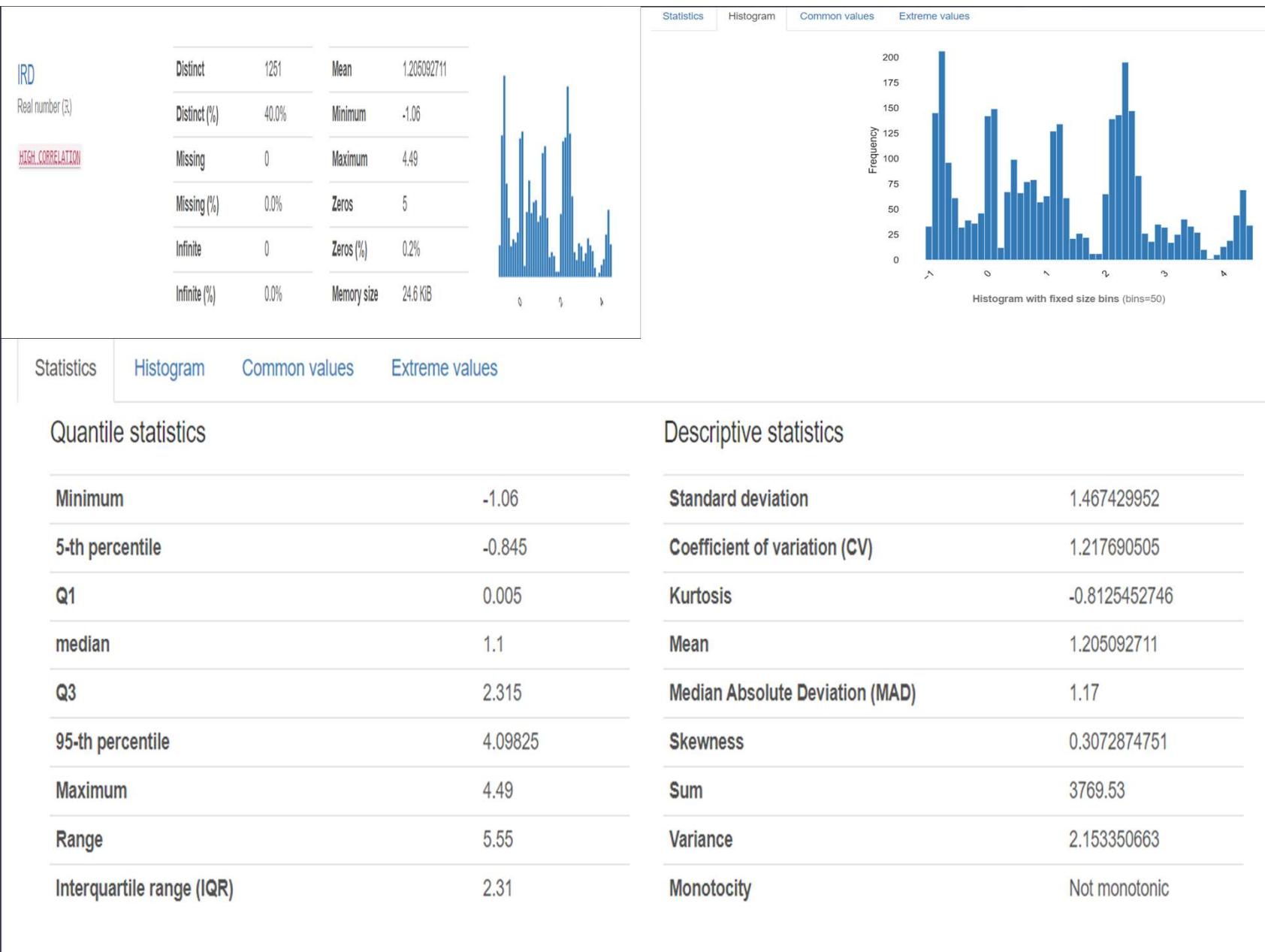
Quantile statistics

Minimum	3863.9
5-th percentile	4199.415
Q1	4978.9
median	5455.55
Q3	5922.325
95-th percentile	6671.395
Maximum	7162.5
Range	3298.6
Interquartile range (IQR)	943.425

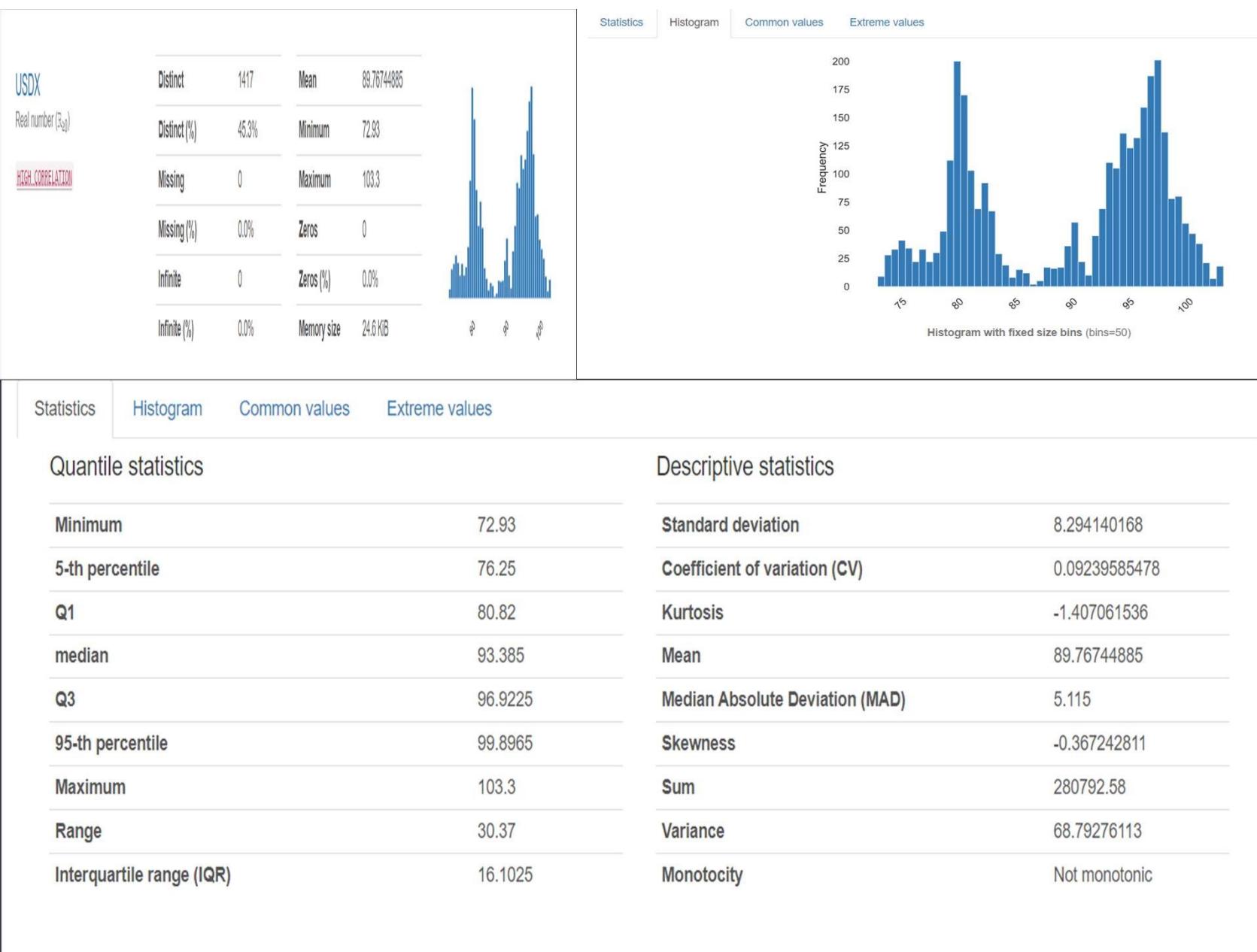
Descriptive statistics

Standard deviation	710.2605037
Coefficient of variation (CV)	0.1305786411
Kurtosis	-0.5125103502
Mean	5439.33141
Median Absolute Deviation (MAD)	471
Skewness	-0.07356210743
Sum	17014228.65
Variance	504469.9831
Monotocity	Not monotonic

IRD STATISTICS AND HISTOGRAM



USDX STATISTICS AND HISTOGRAM



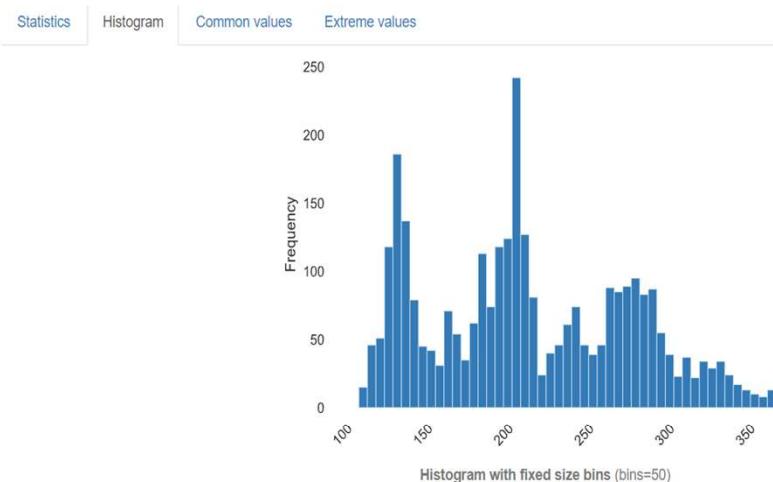
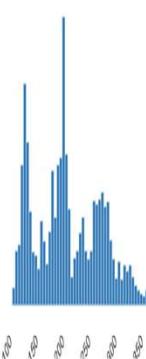
SP 500 STATISTICS AND HISTOGRAM

SP500

Real number (32)

HIGH CORRELATION

Distinct	2338	Mean	216.9149105
Distinct (%)	74.7%	Minimum	109.93
Missing	0	Maximum	373.88
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	24.6 kB



Statistics

Histogram

Common values

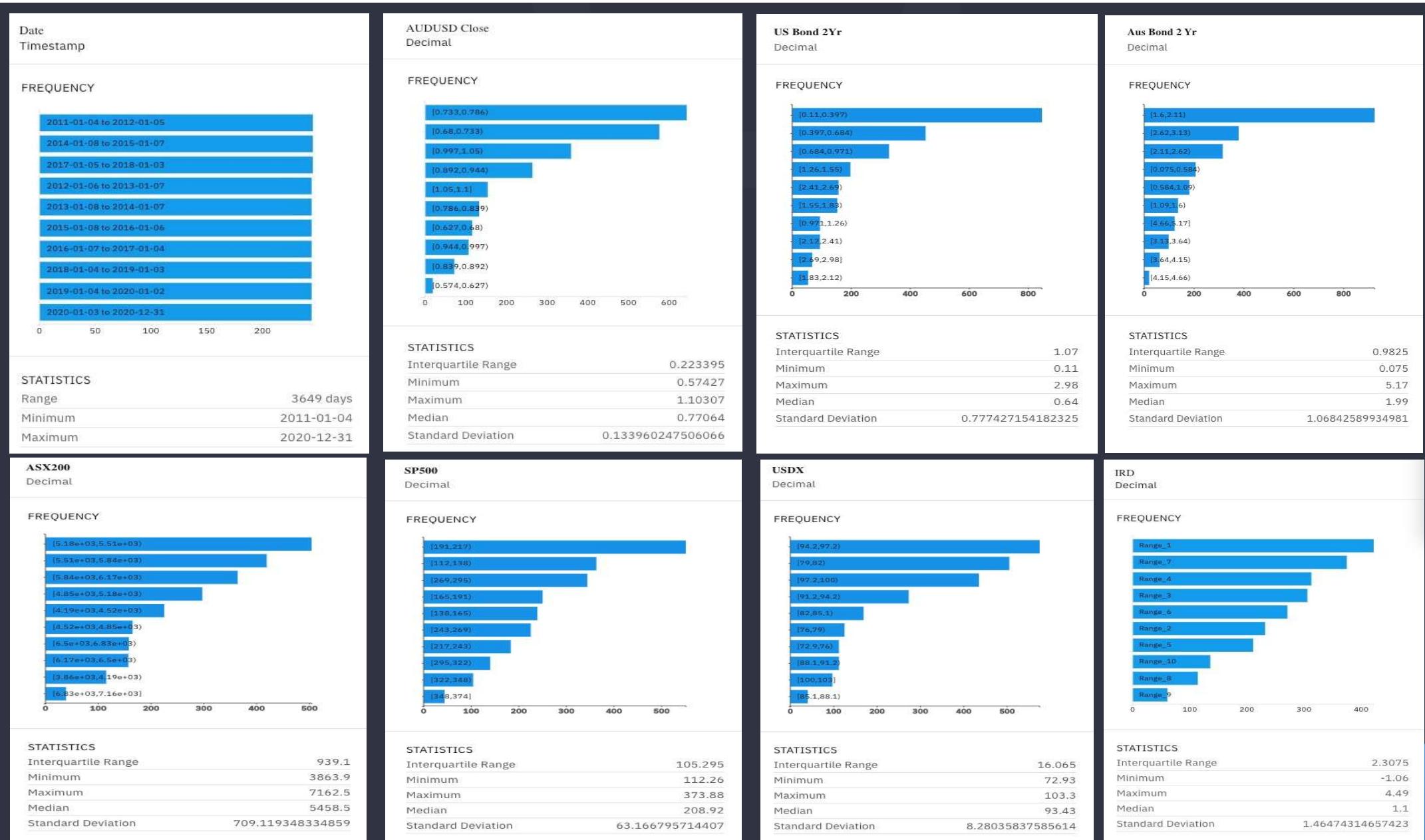
Extreme values

Quantile statistics

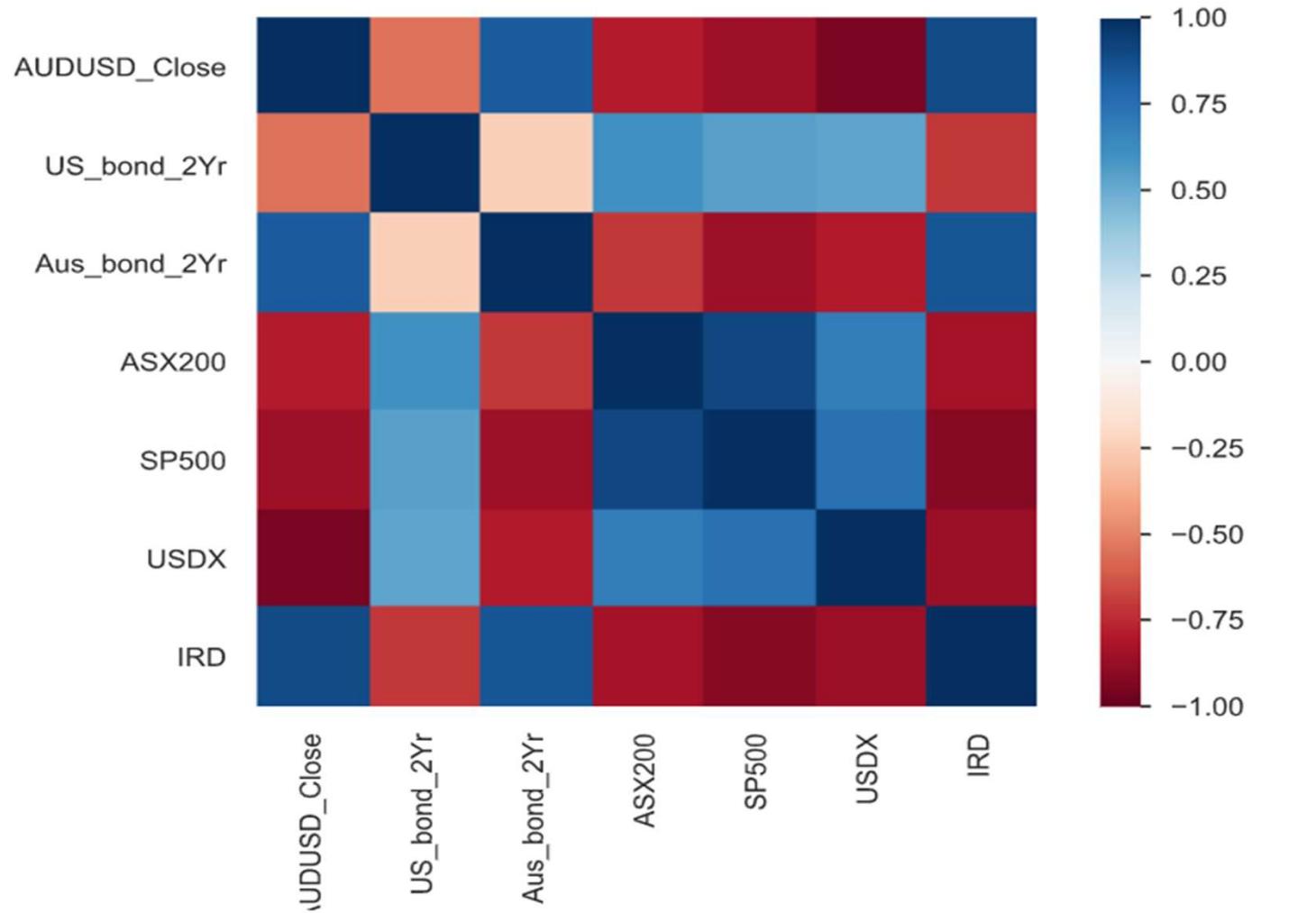
Minimum	109.93
5-th percentile	127.8605
Q1	165.175
median	208.705
Q3	270.5225
95-th percentile	326.79
Maximum	373.88
Range	263.95
Interquartile range (IQR)	105.3475

Descriptive statistics

Standard deviation	63.31855856
Coefficient of variation (CV)	0.2919050535
Kurtosis	-0.8360917556
Mean	216.9149105
Median Absolute Deviation (MAD)	55.055
Skewness	0.27988456
Sum	678509.84
Variance	4009.239858
Monotonicity	Not monotonic



HEATMAP



First rows

	Date	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
0	2011-01-02	1.02120	0.61	5.100	4742.5	127.05	79.13	4.490
1	2011-01-03	1.01665	0.61	5.100	4742.5	127.05	79.13	4.490
2	2011-01-04	1.00691	0.63	5.100	4742.5	126.98	79.44	4.470
3	2011-01-05	0.99974	0.71	5.030	4714.9	127.64	80.26	4.320
4	2011-01-06	0.99287	0.68	5.040	4725.0	127.39	80.79	4.360
5	2011-01-07	0.99558	0.60	5.030	4705.0	127.14	81.01	4.430
6	2011-01-09	0.99512	0.60	5.030	4705.0	127.14	81.01	4.430
7	2011-01-10	0.99465	0.59	4.985	4712.3	126.98	80.88	4.395
8	2011-01-11	0.98218	0.60	4.840	4710.7	127.43	80.85	4.240
9	2011-01-12	0.99571	0.61	4.890	4724.2	128.58	80.03	4.280

Last rows

	Date	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
3118	2020-12-21	0.75753	0.13	0.090	6669.9	367.86	90.04	-0.040
3119	2020-12-22	0.75536	0.13	0.085	6599.6	367.24	90.65	-0.045
3120	2020-12-23	0.75822	0.13	0.085	6643.1	367.57	90.41	-0.045
3121	2020-12-24	0.75919	0.13	0.080	6664.8	369.00	90.32	-0.050
3122	2020-12-25	0.76040	0.13	0.080	6664.8	369.00	90.32	-0.050
3123	2020-12-27	0.76100	0.13	0.080	6664.8	369.00	90.32	-0.050
3124	2020-12-28	0.75871	0.13	0.080	6664.8	372.17	90.34	-0.050
3125	2020-12-29	0.76097	0.12	0.075	6700.3	371.46	89.99	-0.045
3126	2020-12-30	0.76874	0.12	0.080	6682.4	371.99	89.68	-0.040
3127	2020-12-31	0.76928	0.13	0.075	6587.1	373.88	89.94	-0.055

Sample of the first 10 data points and the last 10 data points

Covariance

is a statistical tool used to determine the relationship of movement between two variables. In finance, the variables can be price returns of different assets. A positive covariance means two asset prices tend to move in the same direction. For example, when the European economy was recovering, Euros and British Pounds appreciated in value simultaneously. A negative covariance means asset prices tend to move in the opposite direction. For example, when the economy was slowing down, people took money away from the stock market and invested in safer assets like government treasuries. So stock prices were declining while fixed income prices were rising. As you can see in the plot, if variable X and Y have positive covariance, they tend to increase in value together; while with negative covariance, X tends to increase in value when Y decreases in value.

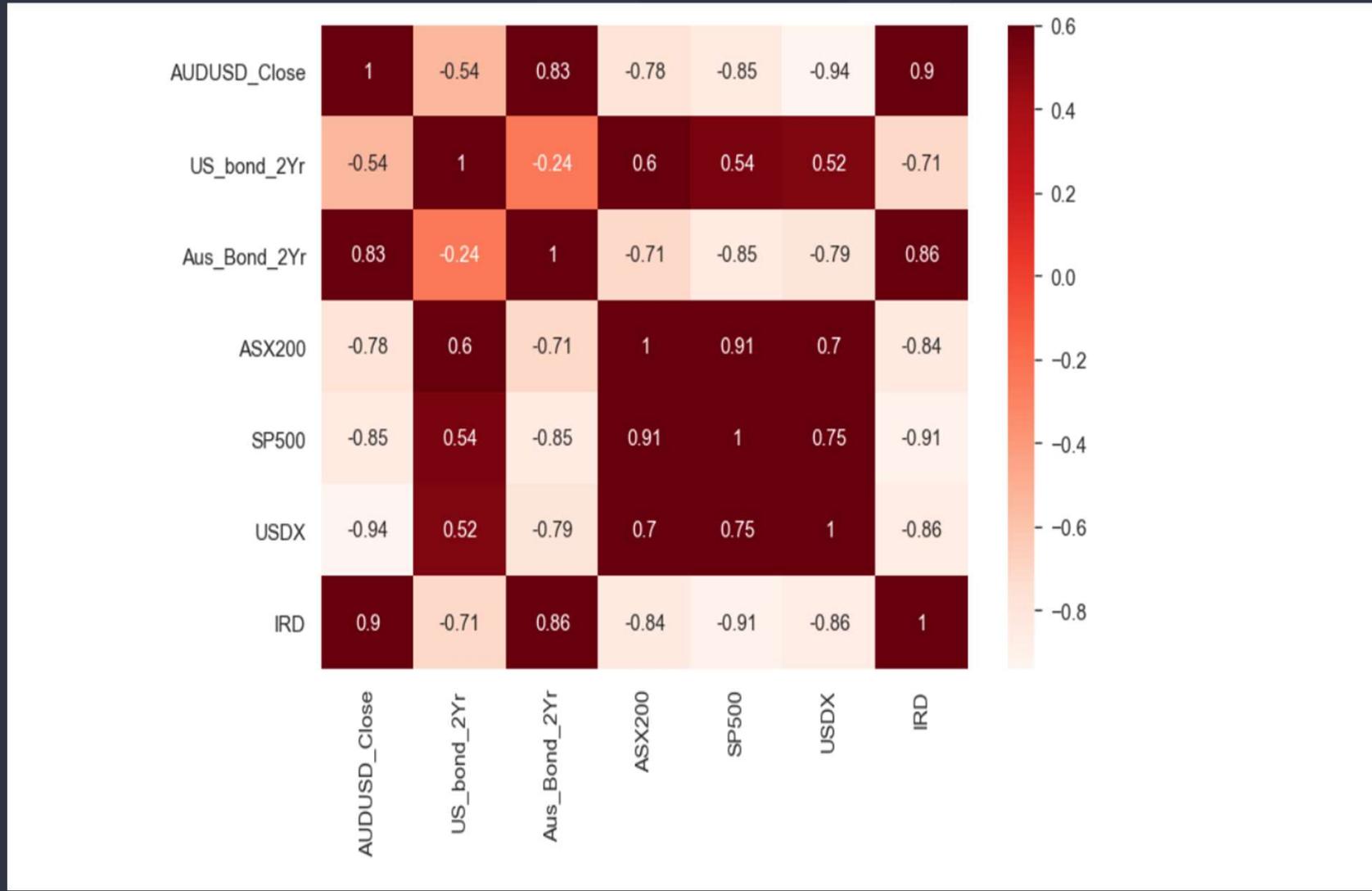
Covariance and Correlation

df.cov()

	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
AUDUSD_Close	0.017994	-0.056444	0.119685	-74.639053	-7.199381	-1.044894	0.176129
US_bond_2Yr	-0.056444	0.604674	-0.201364	332.624378	26.745107	3.377576	-0.806038
Aus_bond_2Yr	0.119685	-0.201364	1.145949	-537.115510	-57.471553	-7.041516	1.347313
ASX200	-74.639053	332.624378	-537.115510	504469.983146	40912.192120	4087.753835	-869.739887
SP500	-7.199381	26.745107	-57.471553	40912.192120	4009.239858	392.266292	-84.216659
USDX	-1.044894	3.377576	-7.041516	4087.753835	392.266292	68.792761	-10.419092
IRD	0.176129	-0.806038	1.347313	-869.739887	-84.216659	-10.419092	2.153351

df.corr()

	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD
AUDUSD_Close	1.000000	-0.541120	0.833480	-0.783405	-0.847621	-0.939157	0.894769
US_bond_2Yr	-0.541120	1.000000	-0.241902	0.602249	0.543191	0.523689	-0.706379
Aus_bond_2Yr	0.833480	-0.241902	1.000000	-0.706427	-0.847890	-0.793071	0.857686
ASX200	-0.783405	0.602249	-0.706427	1.000000	0.909712	0.693898	-0.834477
SP500	-0.847621	0.543191	-0.847890	0.909712	1.000000	0.746928	-0.906379
USDX	-0.939157	0.523689	-0.793071	0.693898	0.746928	1.000000	-0.856054
IRD	0.894769	-0.706379	0.857686	-0.834477	-0.906379	-0.856054	1.000000



05

Train Data Models

Predictive Analytics

20%

40%

60%

80%

100%

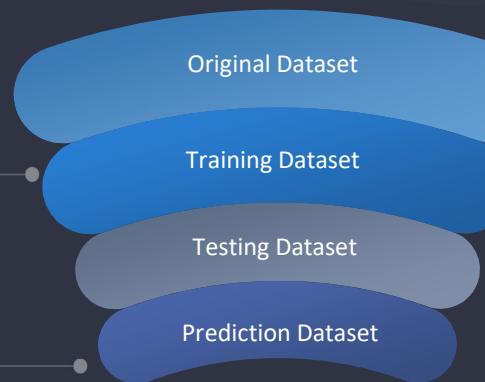
TRAINING VS. TESTING VS PREDICTION DATASET

Training Dataset

9 years, 2814 records
From Jan 1, 2011 till Dec. 31 2019

Prediction Dataset

3.5 months, 89 records
From Jan 1, 2021 till April 15 2021



Original Dataset

13 years++, 4467 records
From Jan 1, 2007 till Apr. 15, 2021

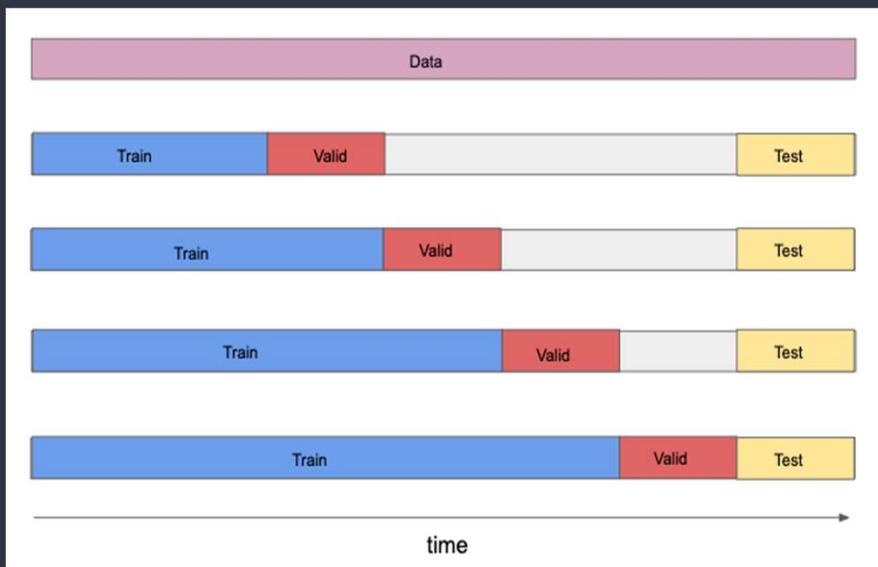
Testing Dataset

1 years, 314 records
From Jan 1, 2020 till Dec. 31 2020

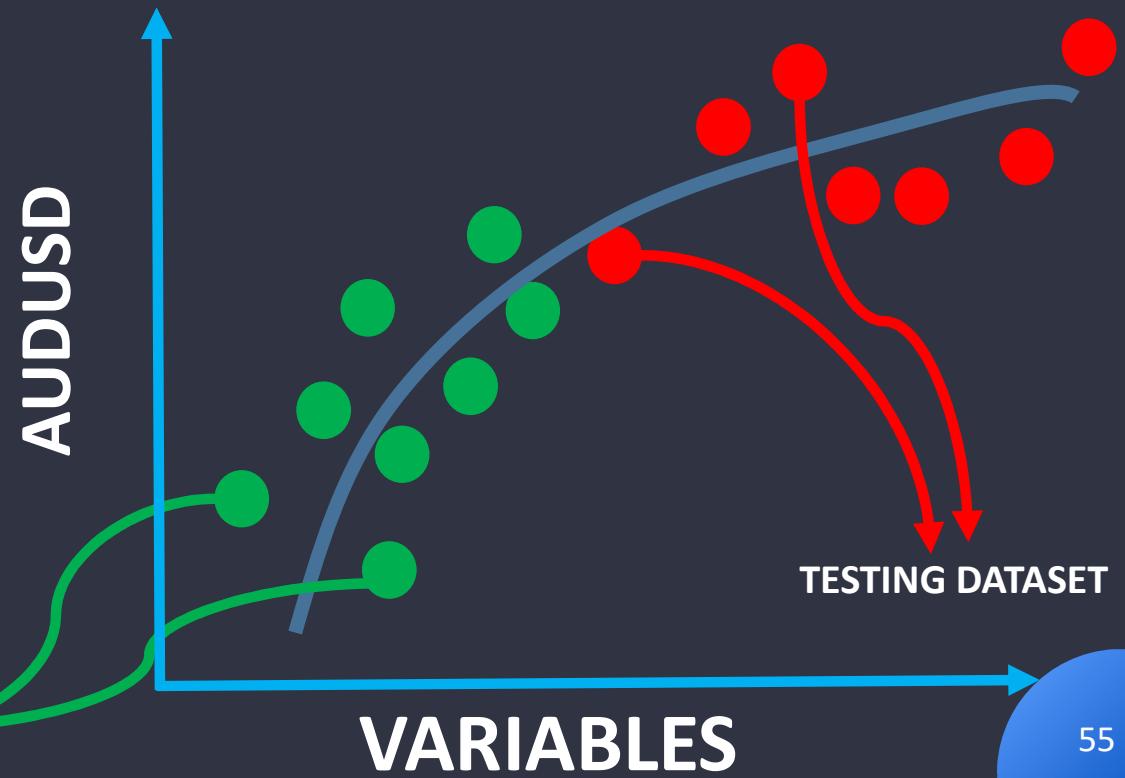


TRAINING VS. TESTING DATASETS

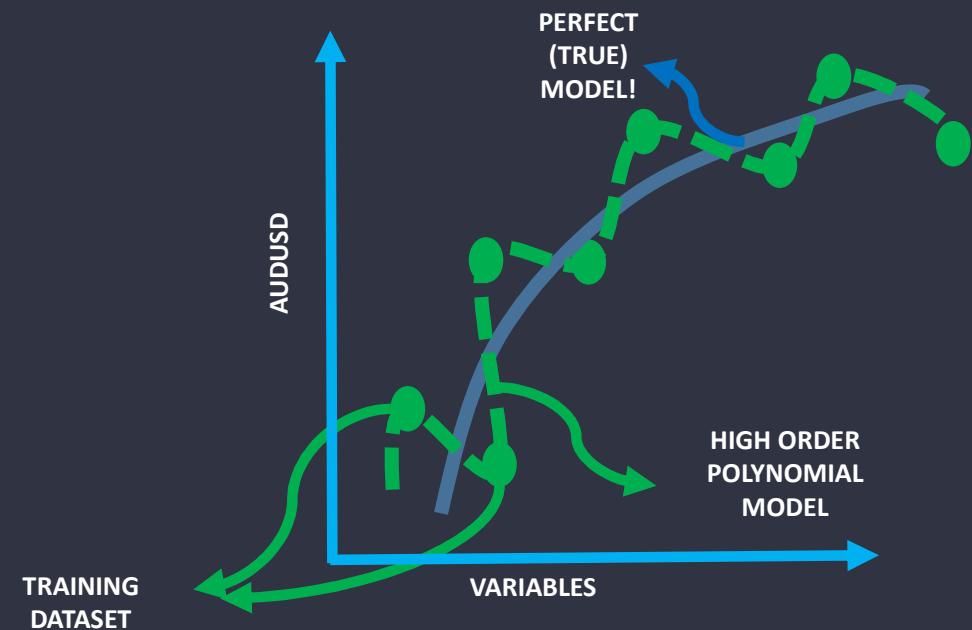
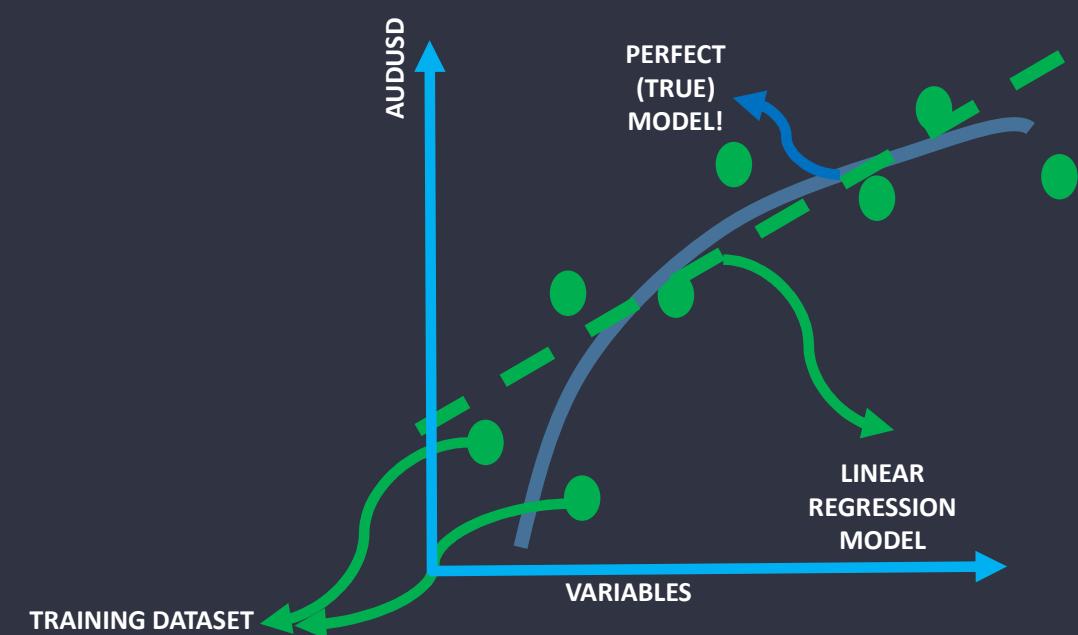
- Dataset is divided to training and testing datasets
- Testing datasets have never been seen by the model before



TRAINING DATASET



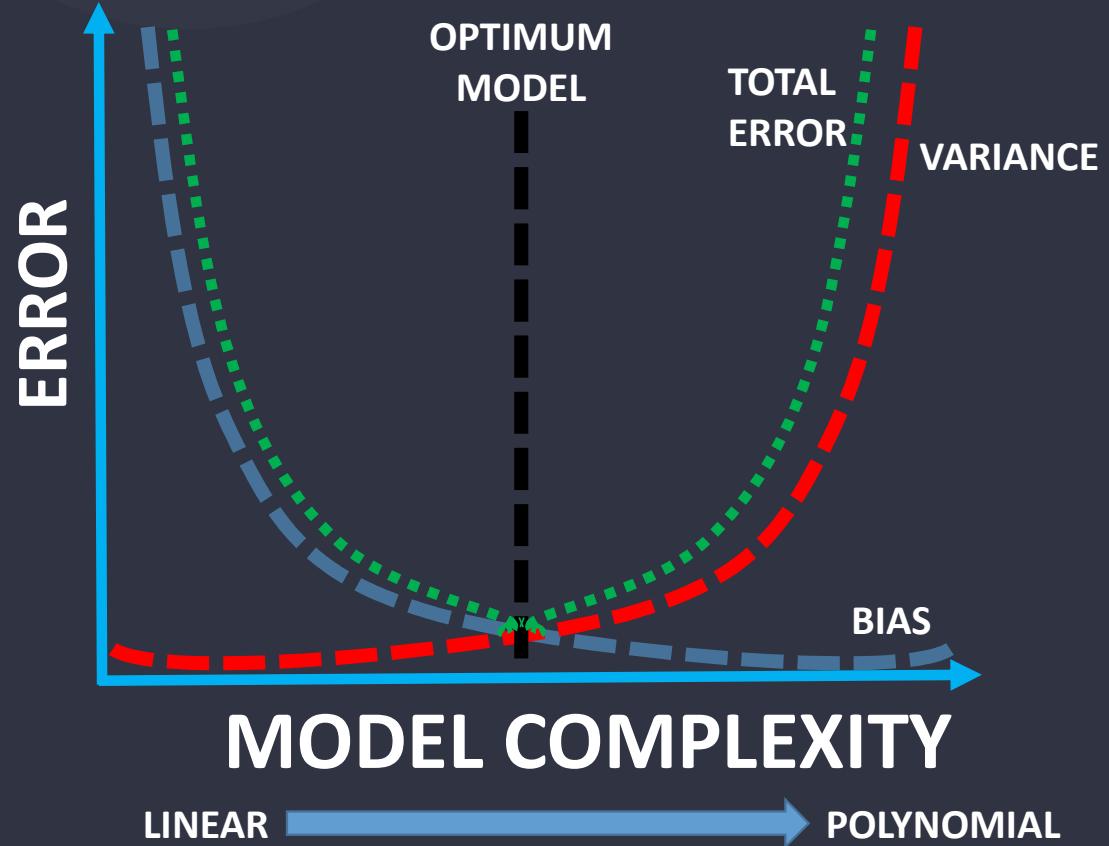
LINEAR REGRESSION (SIMPLE/MULTIPLE)



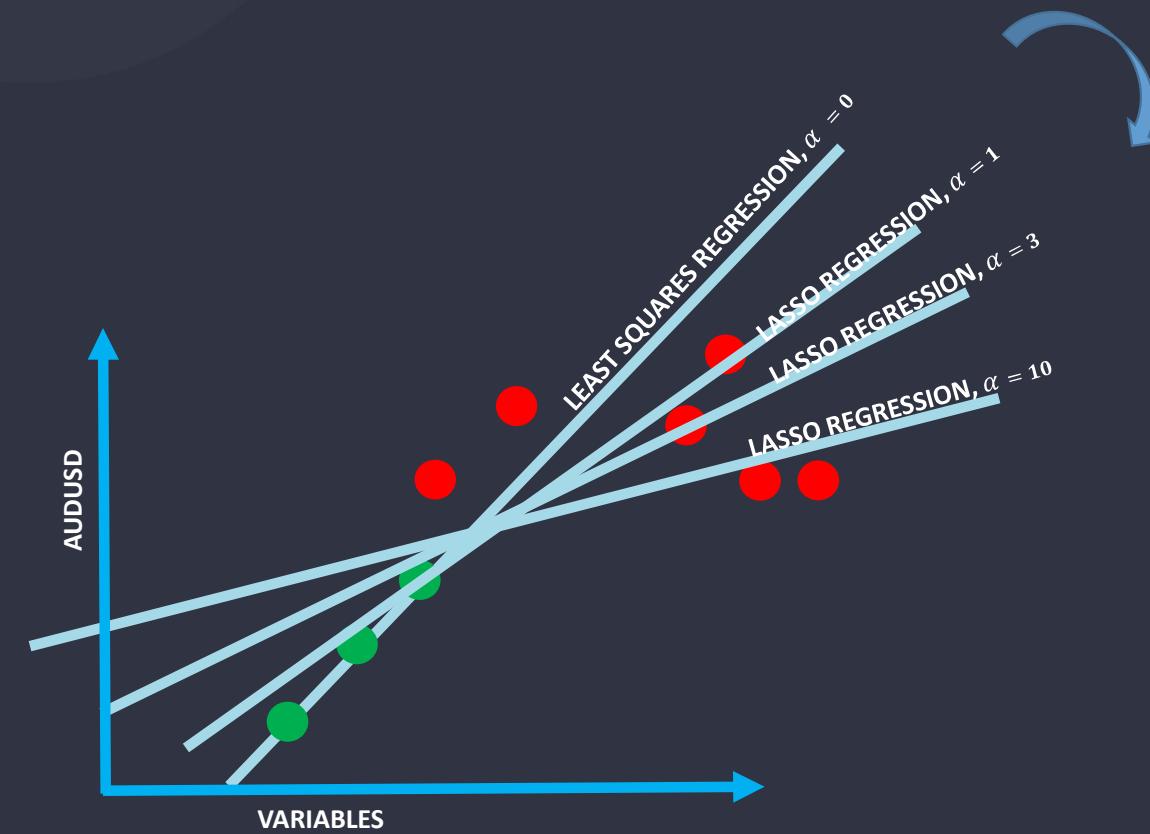
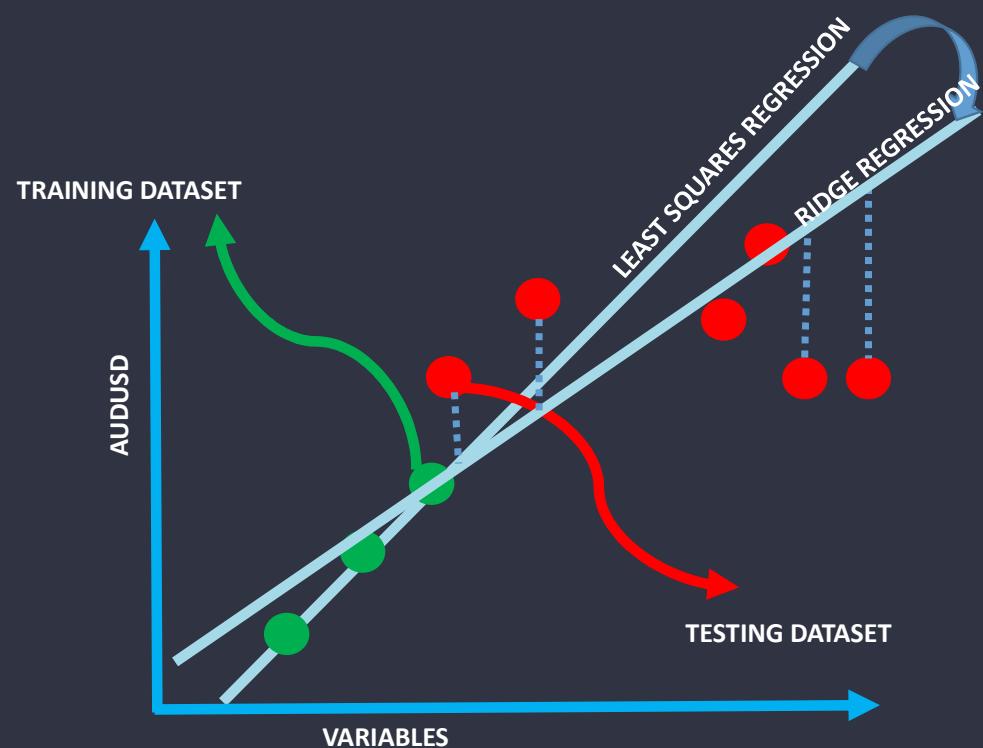
POLYNOMIAL REGRESSION (COMPLEX)

MODEL COMPLEXITY VS. ERROR

- Regularization works by reducing the variance at the cost of adding some bias to the model.
- A trade-off between variance and bias occurs.



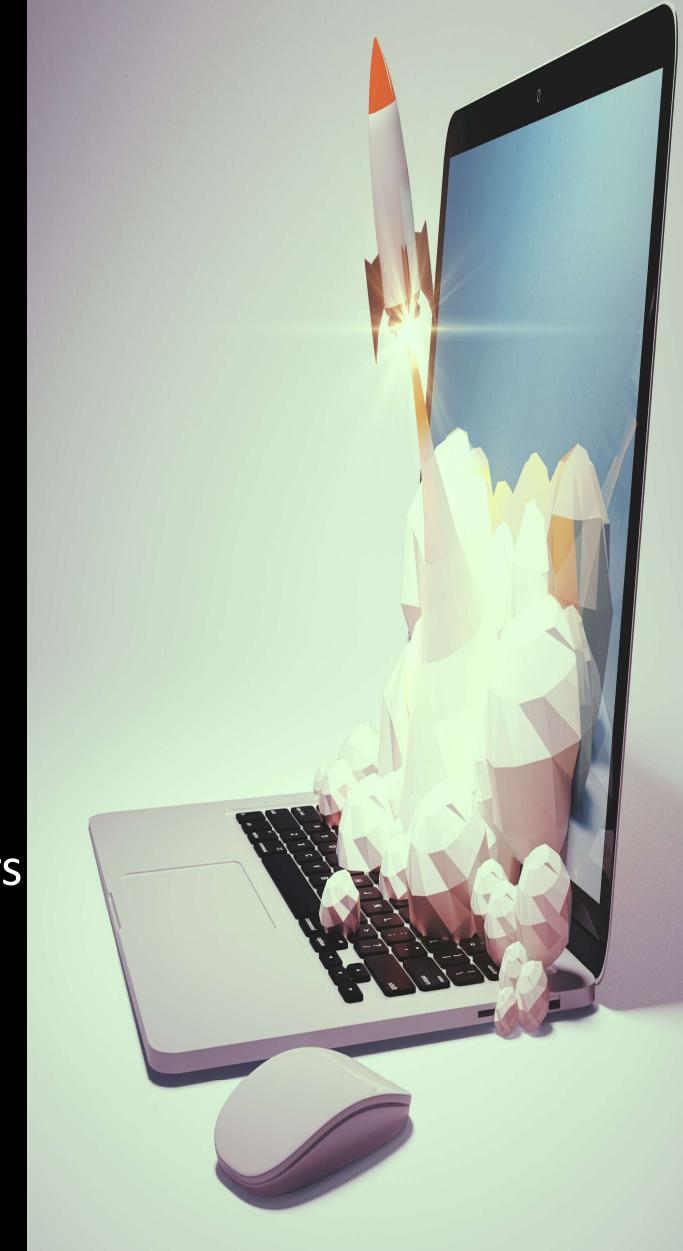
RIDGE REGRESSION (L2 REGULARIZATION)



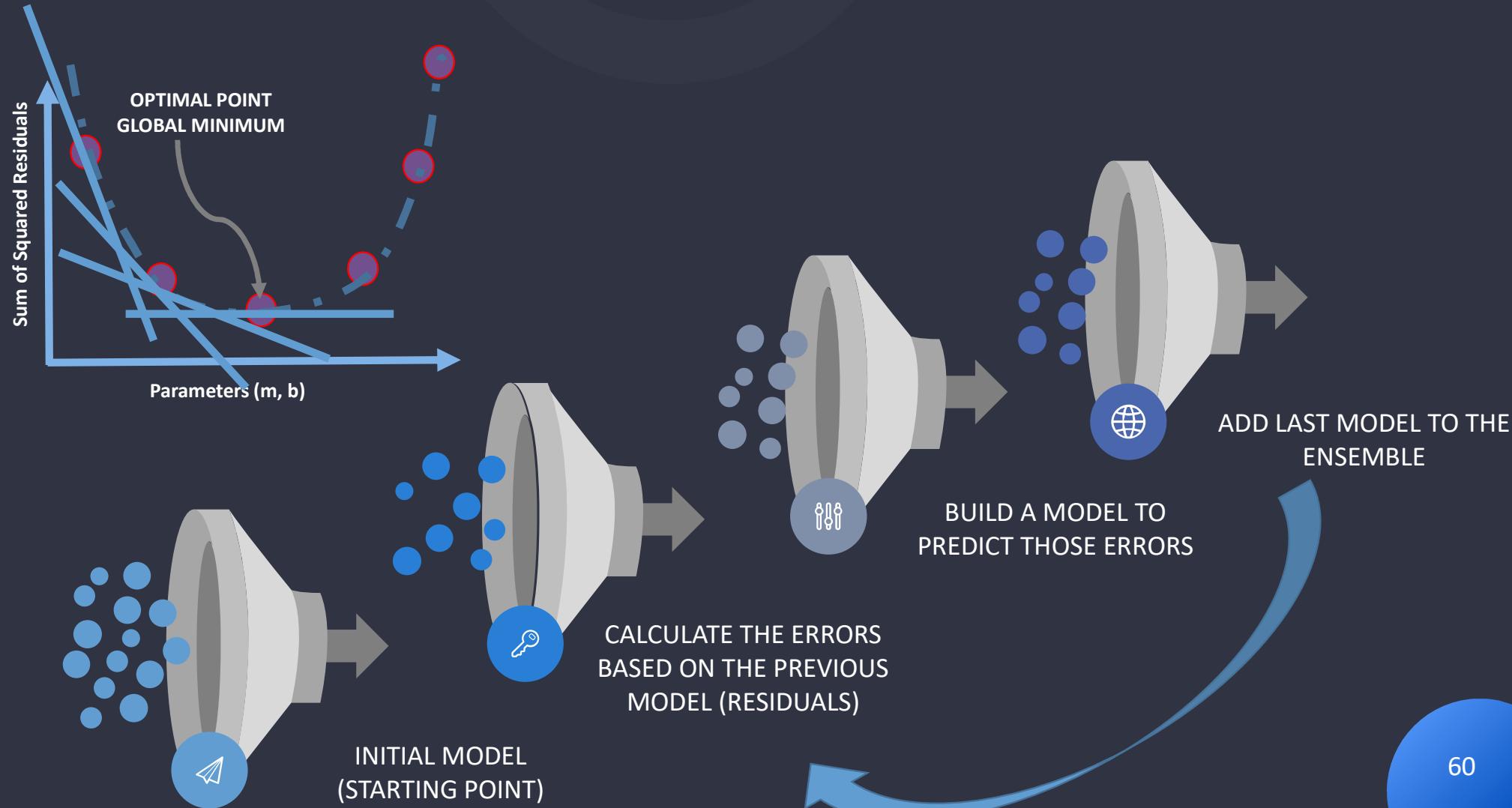
LASSO REGRESSION (L1 REGULARIZATION)

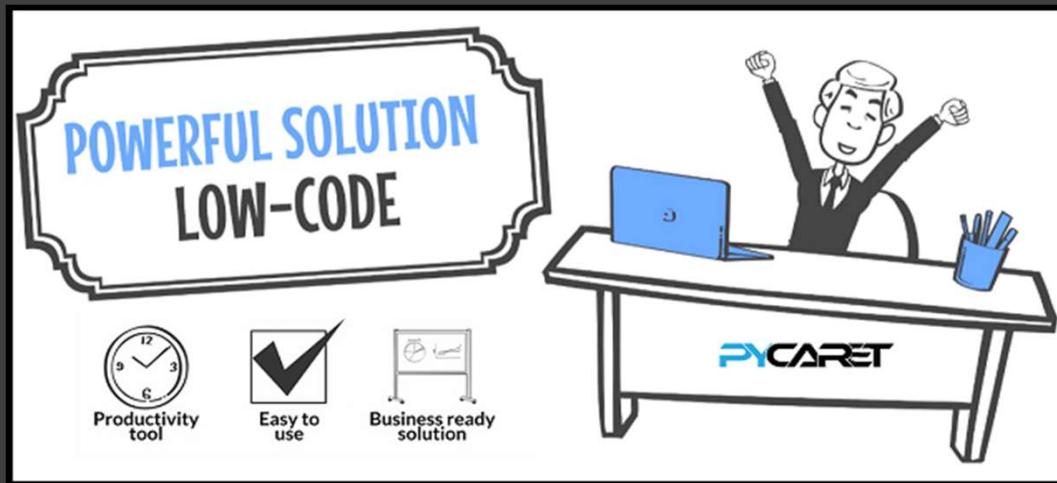
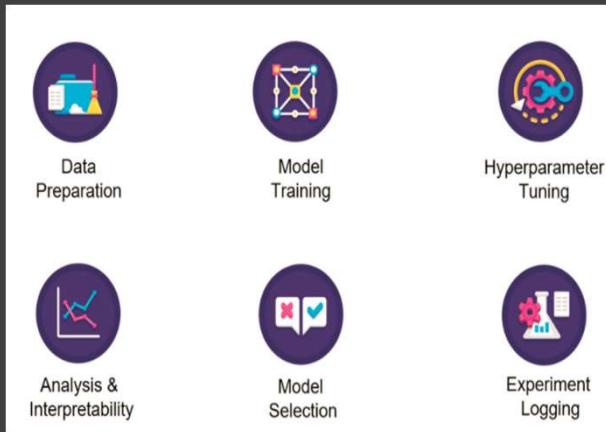
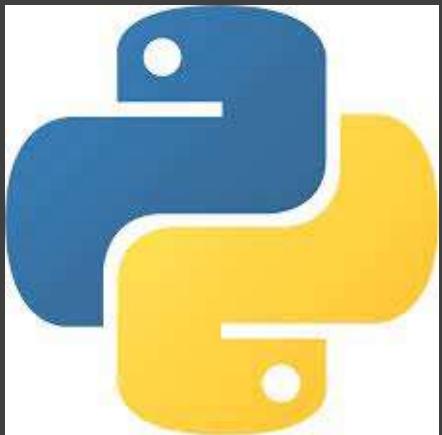
XGBOOST OR EXTREME GRADIENT BOOSTING: INTRODUCTION

- algorithm of choice for regression and classification tasks.
- supervised learning algorithm and implements gradient boosted trees algorithm.
- combining an ensemble of predictions from several weak models.
- robust to many data distributions and relationships and offers many hyperparameters to tune model performance.
- increased speed and enhanced memory utilization.
- “discovering truth by building on previous discoveries”.



XGBOOST: WHAT IS BOOSTING?





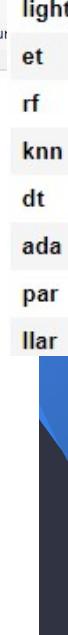
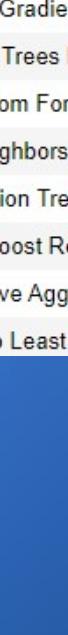
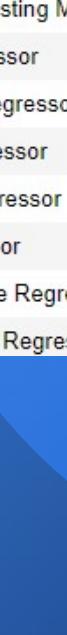
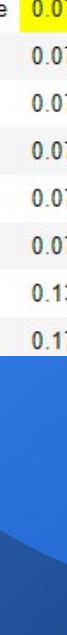
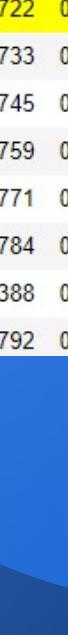
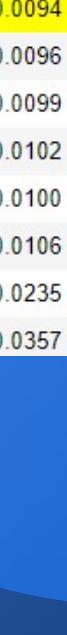
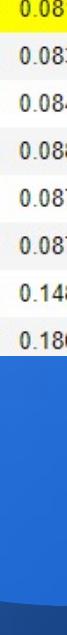
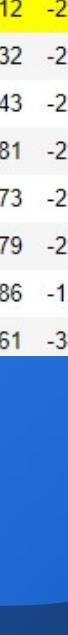
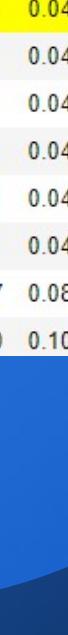
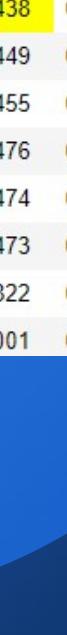
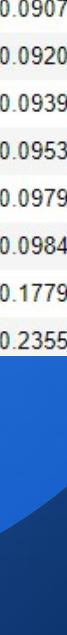
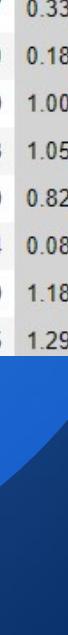
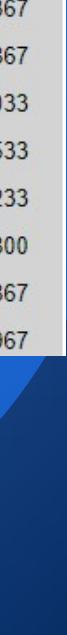
Why Python & PyCaret

- Python has thousands of free libraries, sample codes and tools
- PyCaret has same algo but less code
- Auto selection of models for your datasets
- Fast and no need for other computing power

Series	Year	Month	AUDUSD_Close	US_bond_2Yr	Aus_bond_2Yr	ASX200	SP500	USDX	IRD	
0	1	2011	1	1.02120	0.61	5.100	4742.5	127.05	79.13	4.490
1	2	2011	1	1.01665	0.61	5.100	4742.5	127.05	79.13	4.490
2	3	2011	1	1.00691	0.63	5.100	4742.5	126.98	79.44	4.470
3	4	2011	1	0.99974	0.71	5.030	4714.9	127.64	80.26	4.320
4	5	2011	1	0.99287	0.68	5.040	4725.0	127.39	80.79	4.360
...	
3123	3124	2020	12	0.76100	0.13	0.080	6664.8	369.00	90.32	-0.050
3124	3125	2020	12	0.75871	0.13	0.080	6664.8	372.17	90.34	-0.050
3125	3126	2020	12	0.76097	0.12	0.075	6700.3	371.46	89.99	-0.045
3126	3127	2020	12	0.76874	0.12	0.080	6682.4	371.99	89.68	-0.040
3127	3128	2020	12	0.76928	0.13	0.075	6587.1	373.88	89.94	-0.055

3128 rows × 10 columns

1. Setting up Environment in PyCaret

```
from pycaret.regression import *  
  
# Load Data  
#   
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
#   
  
# 
```

2. Train and Evaluate all Models

```
best = compare_models(sort='MAE')
```

Model		MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
gbr	Gradient Boosting Regressor	0.0722	0.0094	0.0813	-2.0559	0.0439	0.0905	0.1633
lightgbm	Light Gradient Boosting Machine	0.0722	0.0094	0.0812	-2.0213	0.0438	0.0907	0.3367
et	Extra Trees Regressor	0.0733	0.0096	0.0832	-2.2385	0.0449	0.0920	0.1867
rf	Random Forest Regressor	0.0745	0.0099	0.0843	-2.2961	0.0455	0.0939	1.0033
knn	K Neighbors Regressor	0.0759	0.0102	0.0881	-2.9383	0.0476	0.0953	1.0533
dt	Decision Tree Regressor	0.0771	0.0100	0.0873	-2.7340	0.0474	0.0979	0.8233
ada	AdaBoost Regressor	0.0784	0.0106	0.0879	-2.9375	0.0473	0.0984	0.0800
par	Passive Aggressive Regressor	0.1388	0.0235	0.1486	-12.7557	0.0822	0.1779	1.1867
llar	Lasso Least Angle Regression	0.1792	0.0357	0.1861	-35.9590	0.1001	0.2355	1.2967

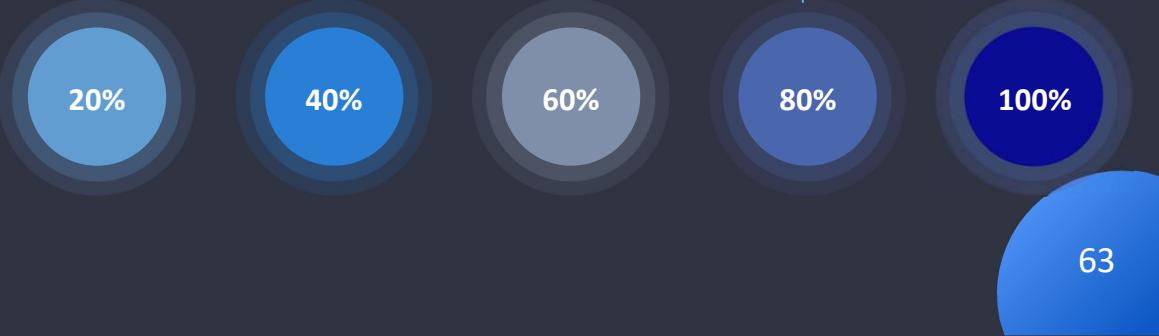
```
15    CPU Jobs      -1  
16    Use GPU       False
```

Validate Data Models



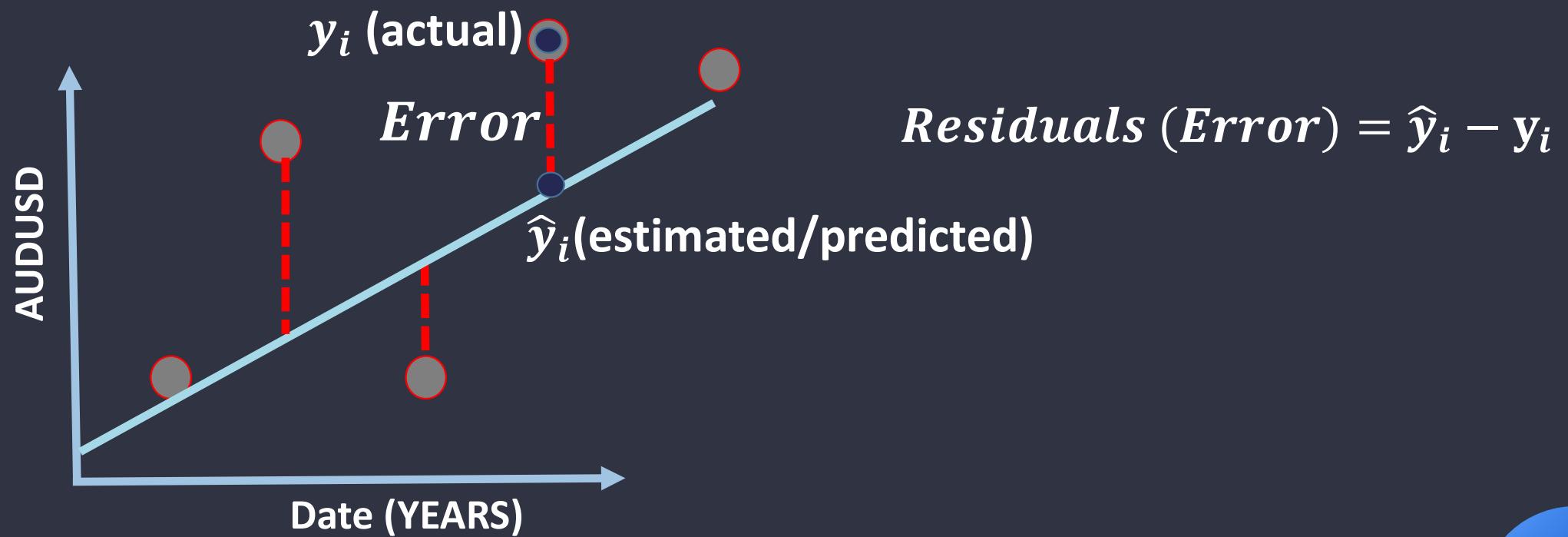
06

Predictive Analytics



REGRESSION METRICS: HOW TO ASSESS MODEL PERFORMANCE?

- After model fitting, we would like to assess the performance of the model by comparing model predictions to actual (True) data



REGRESSION METRICS: MEAN ABSOLUTE ERROR (MAE)

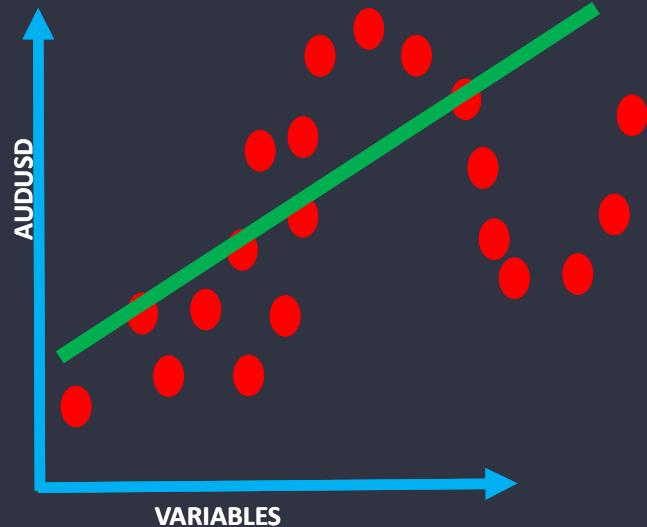
- obtained by calculating the absolute difference between the model predictions and the true (actual) values
- MAE is a measure of the **average magnitude of error** generated by the regression model

Model		MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
gbr	Gradient Boosting Regressor	0.0722	0.0094	0.0813	-2.0559	0.0439	0.0905	0.1867
lightgbm	Light Gradient Boosting Machine	0.0722	0.0094	0.0812	-2.0213	0.0438	0.0907	0.3200
et	Extra Trees Regressor	0.0733	0.0096	0.0832	-2.2385	0.0449	0.0920	0.2300
rf	Random Forest Regressor	0.0745	0.0099	0.0843	-2.2961	0.0455	0.0939	1.1400
knn	K Neighbors Regressor	0.0759	0.0102	0.0881	-2.9383	0.0476	0.0953	1.3533
dt	Decision Tree Regressor	0.0771	0.0100	0.0873	-2.7340	0.0474	0.0979	1.1000
ada	AdaBoost Regressor	0.0784	0.0106	0.0879	-2.9375	0.0473	0.0984	0.1133
par	Passive Aggressive Regressor	0.1388	0.0235	0.1486	-12.7557	0.0822	0.1779	1.3033
llar	Lasso Least Angle Regression	0.1792	0.0357	0.1861	-35.9590	0.1001	0.2355	1.3767

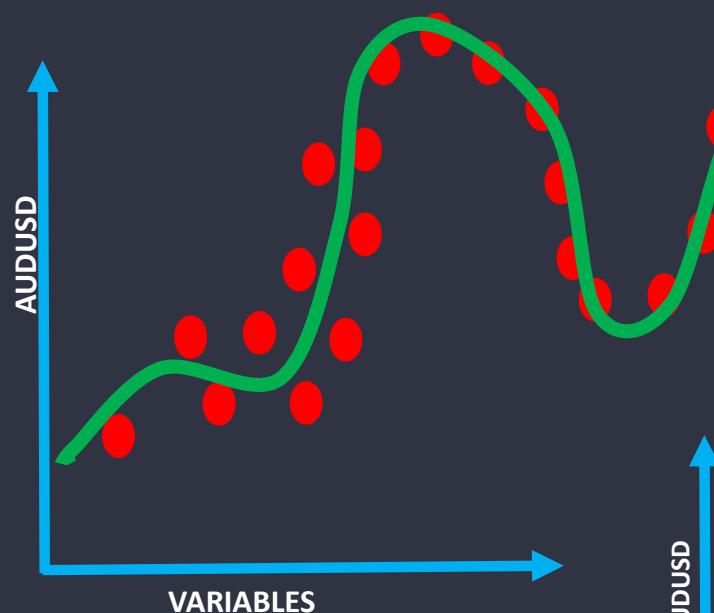


BEST MODEL

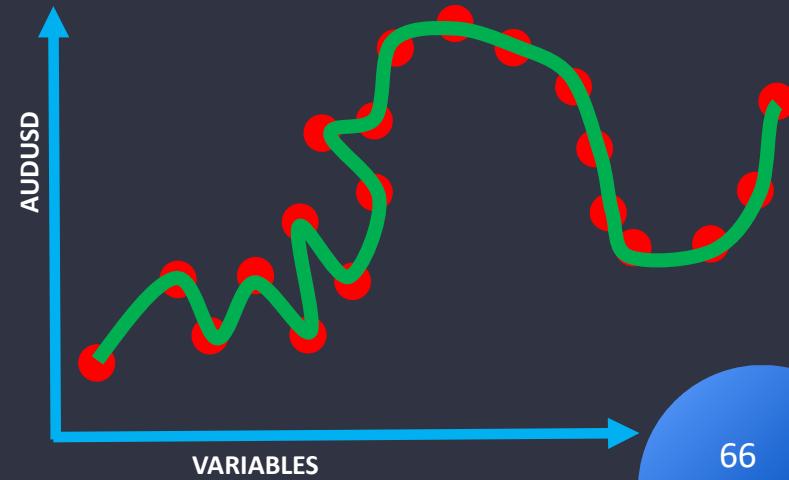
UNDERFITTING
MODEL IS TOO
SIMPLE FOR THIS COMPLEX
DATASET



GENERALIZED
MODEL



MODEL IS OVERFITTING
THE DATA



3. Check the score on the test dataset

```
prediction_holdout = predict_model(best)
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Gradient Boosting Regressor	0.0265	0.0012	0.0348	0.2842	0.0208	0.0395

```
prediction_holdout = predict_model(model)
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Extra Trees Regressor	0.0258	0.0014	0.0373	0.1780	0.0224	0.0392

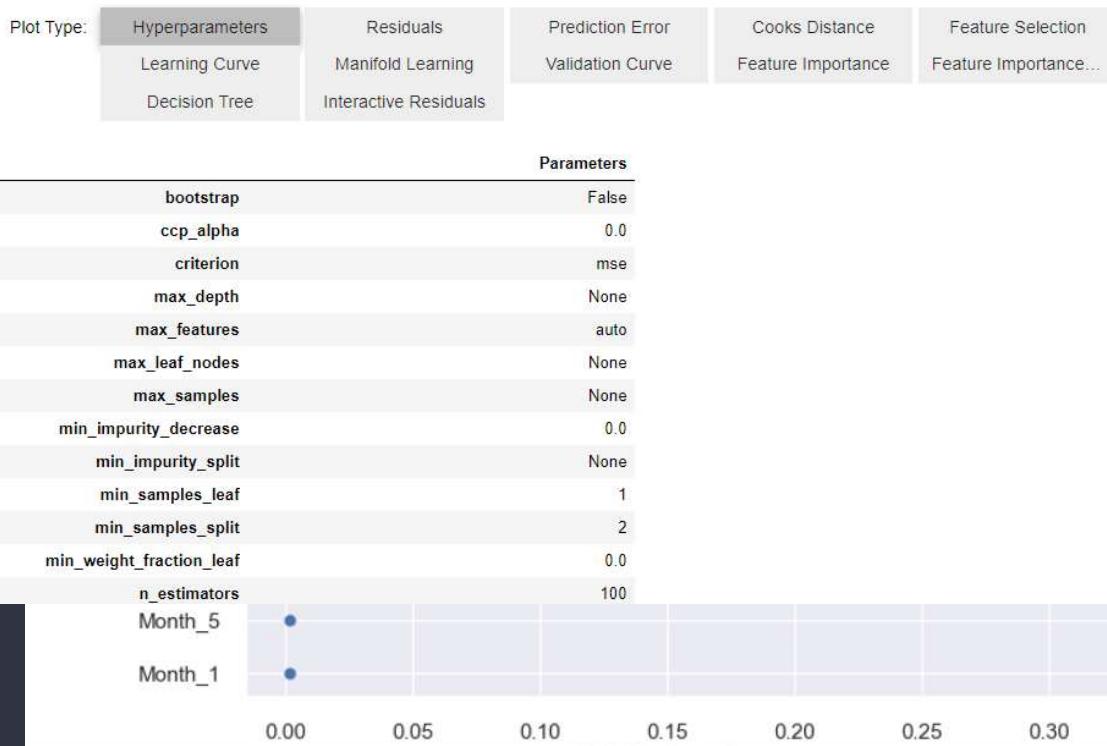
.....ET is 4% better than GBR

4. Examining chosen model

```
model = create_model('et', fold=3)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.1405	0.0243	0.1559	-4.0056	0.0809	0.1652
1	0.0298	0.0015	0.0384	-1.3135	0.0219	0.0410

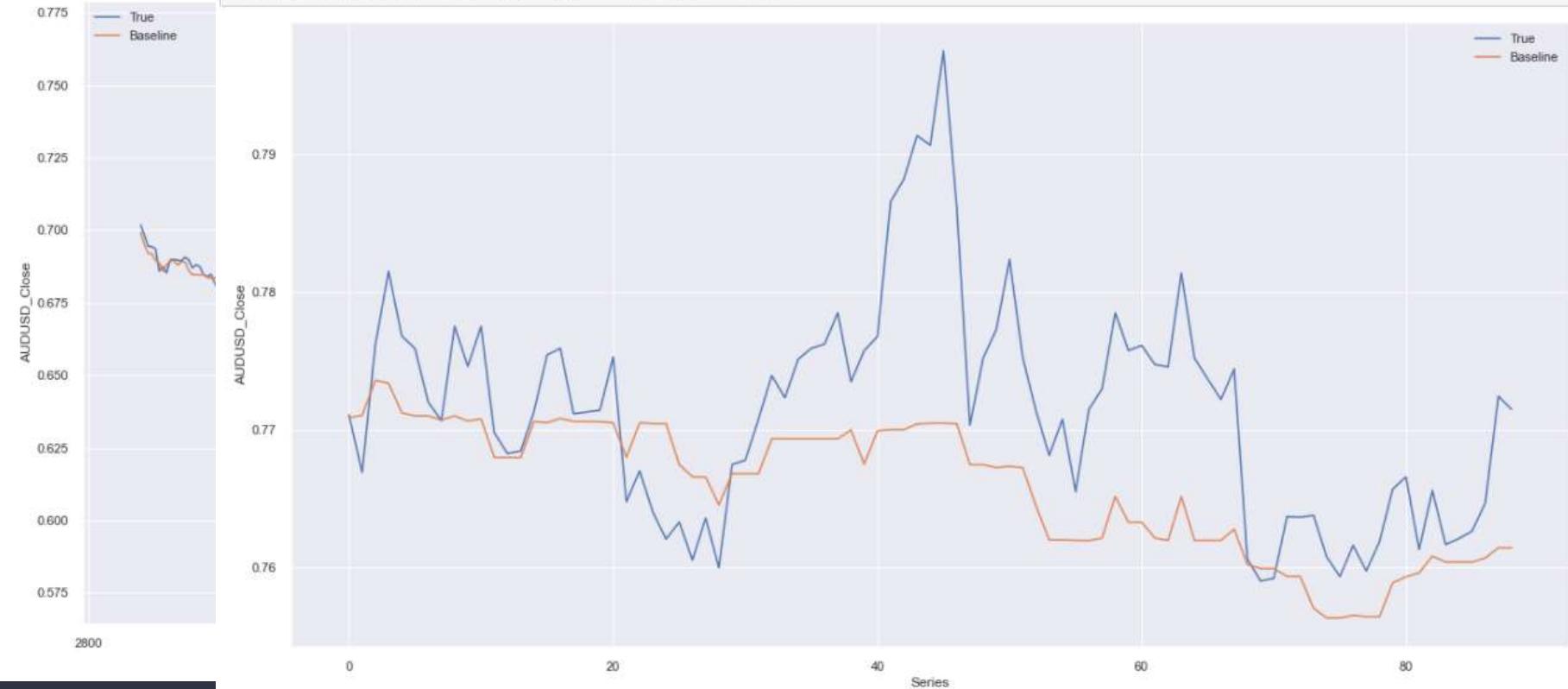
```
evaluate_model(model)
```



5. Plot the actual data and predictions

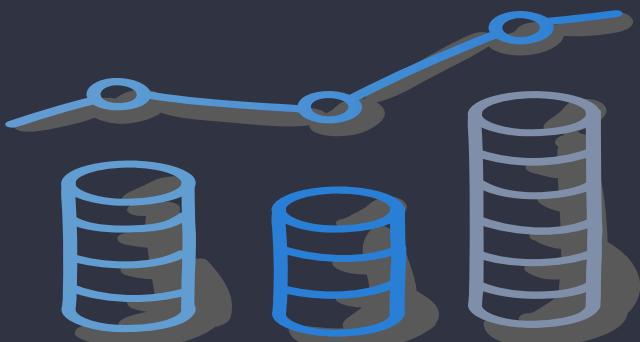
```
plt.figure(figsize=(20,10))
plot_series(audusd)
#plot_series(train['ds'])
plot_series(audusd3)
```

```
plt.figure(figsize=(20,10))
plot_series(audusd3.index, audusd3['AUDUSD_Close'], "True")
#plot_series(train['ds'],train['y'])
plot_series(audusd3.index, predict['Label'], "Baseline")
```

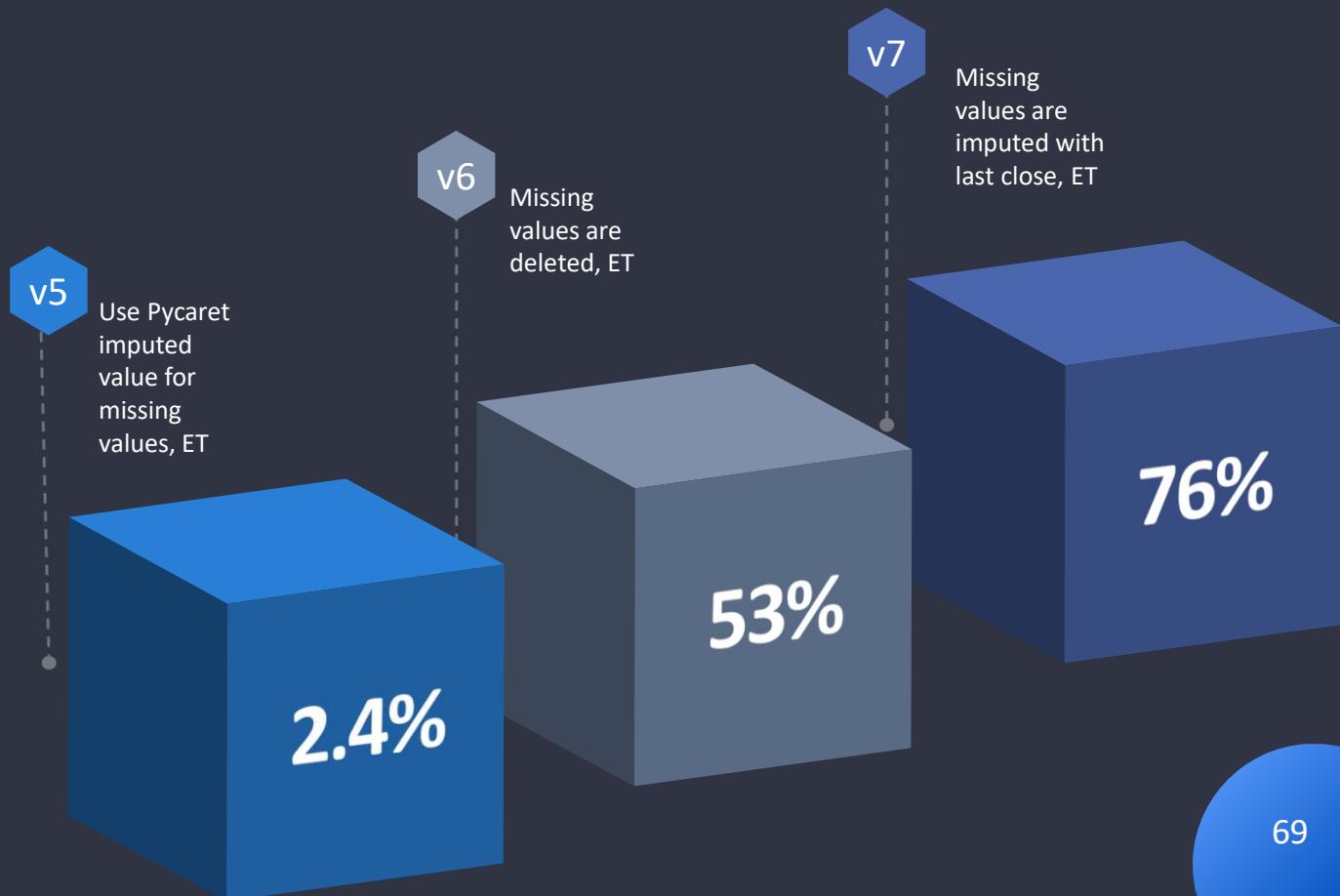


What is Volatility?

- movements in price over time.
- AUDUSD historical daily movement is 100 pips
- Prediction within 100 pips range considered good



Performance of our models

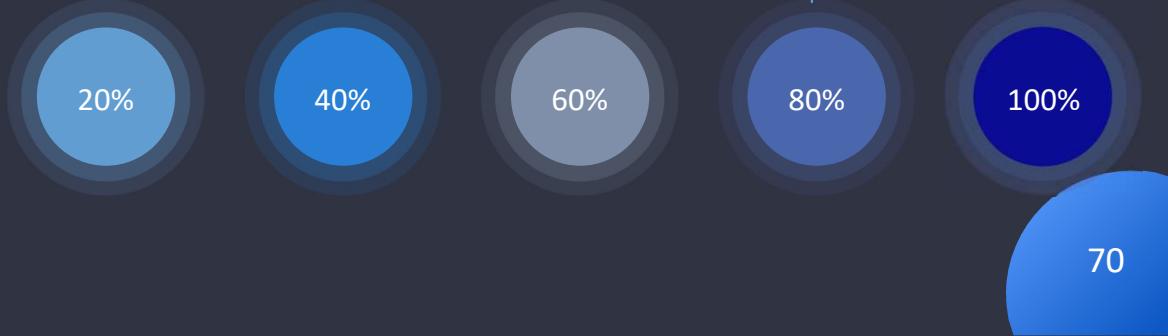


Deploy Data Models



07

Predictive Analytics





Amazon API
Gateway

+



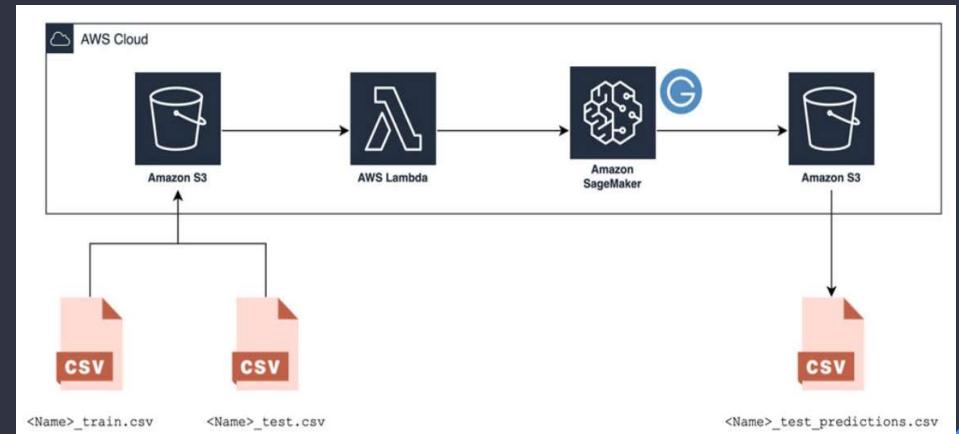
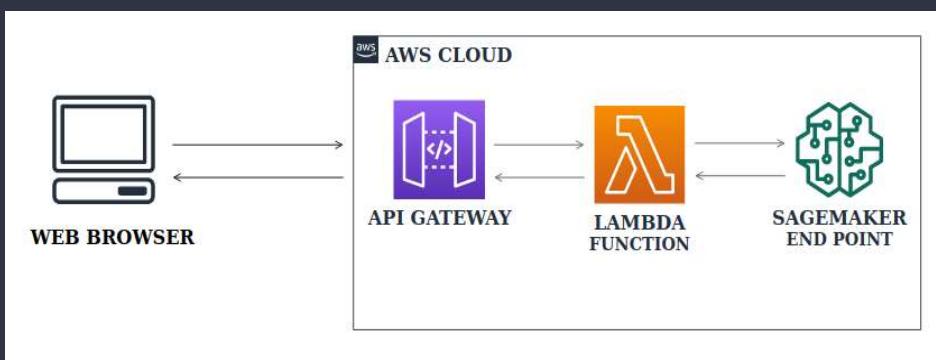
Lambda
function



Amazon SageMaker



Why AWS Lambda and how does it works?



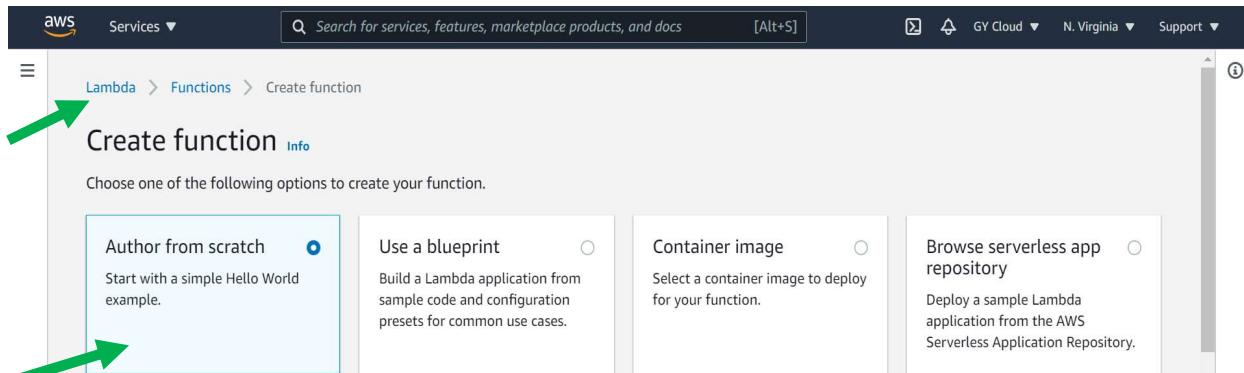
Steps to Deployment

The screenshot shows the AWS SageMaker console interface. On the left, a modal window titled 'Deploy model' is open, showing deployment options for a real-time prediction model. It includes fields for 'Endpoint name' (set to 'GYForexPrediction-endpoint'), 'Instance type' (set to 'ml.m5.xlarge'), 'Instance count' (set to '3'), and a checked 'Save prediction requests' option. On the right, the main SageMaker dashboard displays the 'Endpoints' section. A green arrow points from the 'Endpoints' link in the navigation sidebar to the list of endpoints. The list shows two entries:

Name	ARN	Creation time	Status	Last updated
GYForexPrediction-endpoint1	arn:aws:sagemaker:us-east-1:321506788348:endpoint/gyforexprediction-endpoint1	Apr 23, 2021 10:54 UTC	InService	Apr 23, 2021 11:03 UTC
GYForexPrediction-endpoint	arn:aws:sagemaker:us-east-1:321506788348:endpoint/gyforexprediction-endpoint	Apr 23, 2021 06:51 UTC	InService	Apr 23, 2021 06:59 UTC

• Create an Endpoint for the Model.
• You can view the Endpoint Configurations in SageMaker UI and make the relevant configuration for the endpoint model.

Steps to Deployment



- From the AWS Lambda console, choose to Create function.
- Create a New Role such that Lambda Function has permission to invoke SageMaker endpoint.

A screenshot of the AWS Lambda 'Functions' list page. At the top, there's a navigation bar with the AWS logo, 'Services ▾', a search bar ('Search for services, features, marketplace products, and docs'), and account information ('GY Cloud ▾ N. Virginia ▾ Support ▾'). Below the navigation bar, the path 'Lambda > Functions' is shown. A green arrow points to the 'Create function' button. The main section is titled 'Functions (3)' with the sub-instruction 'Last fetched 2 minutes ago'. There are three functions listed:

Function name	Description	Package type	Runtime	Code size	Last modified
GY-ForexPrediction		Zip	Node.js 14.x	304 bytes	12 hours ago
ForexPrediction		Zip	Python 3.8	299 bytes	2 days ago
gyprediction	An Amazon S3 trigger that retrieves metadata for the object that has been updated.	Zip	Python 3.7	566 bytes	2 days ago

Steps to Deployment

The screenshot shows two screenshots of the AWS API Gateway interface. The top screenshot is a modal titled 'Choose an API type' with 'HTTP API' selected. A green arrow points to the 'HTTP API' section. The bottom screenshot shows the main 'APIs' list with one entry: 'GY-ForexPrediction-API'. A green arrow points to this entry.

• Choose the Create API and select REST API (as client send a request and get the response).

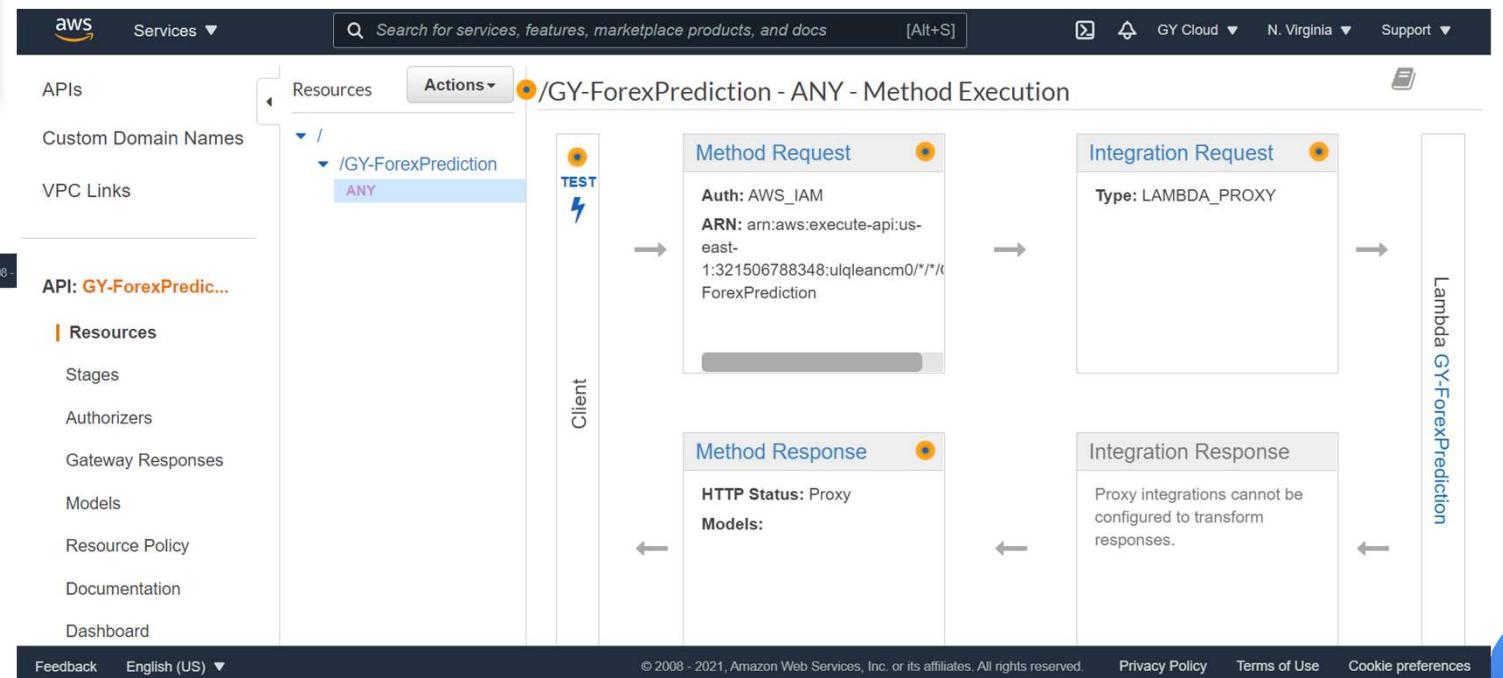
• Give a name to your API and choose endpoint type as Regional.

Name	Description	ID	Protocol	Endpoint type	Created
GY-ForexPrediction-API	Created by AWS Lambda	ulqleancm0	REST	Regional	2021-04-23

Steps to Deployment

The screenshot shows the AWS API Gateway interface. On the left, there's a sidebar with links like APIs, Custom Domain Names, VPC Links, and others. The main area is titled 'GY-ForexPrediction-API' and shows a 'Resources' section. A green arrow points to the 'Actions' dropdown menu, which is open and displays options: 'Create Method', 'Create Resource', 'Enable CORS', and 'Edit Resource Documentation'. Below this, another 'Actions' dropdown is shown with options: 'Deploy API', 'Import API', 'Edit API Documentation', and 'Delete API'.

- Click create resources.
- Following several other steps to configure the integration to Lambda Function with the codes.
- API structure formed.



Steps to Deployment

The screenshot illustrates the deployment process for an API named "GY-ForexPrediction-API".

Top Left: The "Deploy API" dialog is open, prompting the user to choose a stage for deployment. It shows a dropdown for "Deployment stage" set to "[New Stage]" and a text input for "Stage name" containing "GYForexPredictions". Below these fields are "Stage description" and "Deployment description" inputs, both currently empty. At the bottom are "Cancel" and "Deploy" buttons.

Top Right: A separate window titled "Lambda GY-ForexPrediction" is visible, showing an integration request type "LAMBDA_PROXY".

Middle: The main API Gateway interface shows the deployed stage "GYForexPredictions". A green arrow points from the "Invoke URL" field in the Stage Editor to the "Default Method Throttling" section below it.

Bottom: The Stage Editor details the "GYForexPredictions" stage, which includes tabs for Settings, Logs/Tracing, Stage Variables, SDK Generation, Export, and Deployment History. The "Documentation History" tab is selected. The "Cache Settings" section contains an "Enable API cache" checkbox. The "Default Method Throttling" section specifies a rate of 10000 requests per second and a burst of 5000 requests. The "Enable throttling" checkbox is checked.

- From Actions menu, select Deploy API.
- On the page that appears, create a new stage. Give it a name and click on Deploy.

Deployment in AutoAI – Time Series

The screenshot shows the AutoAI platform interface for a project titled "AUDUSD_Prediction-V7ForTimeSeries".

Left Panel: Shows the "Add data sources" section with a "Merged Data_Edited_V7.csv" file selected. A green arrow points to the "PREDICTION MODE" dropdown menu, which is currently set to "Time Series".

Top Center: A modal window titled "Configure details" asks "Create a time series forecast?" with a checked "Yes" option and a "No" button.

Middle Center: A modal window titled "Discarded pipelines" lists seven pipelines that were considered but discarded:

Name	Algorithm	Projected SMAPE	Training data used
Pipeline 1	Random Forest	0.518	79%
Pipeline 10	BATS	0.407	79%
Pipeline 2	SVM	0.412	79%
Pipeline 3	Linear Regression	0.412	79%
Pipeline 7	Holt-Winters	0.406	79%
Pipeline 8	Holt-Winters	0.406	79%
Pipeline 9	ARIMA	0.406	79%

Bottom Center: A "View discarded pipelines (7)" button is highlighted with a red circle.

Right Side: A progress map and experiment summary table are visible.

Enhancements	Build time
HPO PE	00:00:01
HPO PE	00:00:06
HPO PE	00:00:01

Testing with AutoAI – Time Series

The screenshot shows the AutoAI interface with the following details:

- Deployment Path:** Deployments / GY Forex Space / Prediction-V7
- Deployment Name:** Prediction-V7
- Scheduled to run:** No schedule created
- Associated Asset:** GY Forex Prediction-V7
- Environment definition:** 2 CPU and 8 GB RAM
- Deployment Job Definition ID:** 5600c51b-b4f9-47f1-86cc-503e9d7f4f70
- Job ID:** 02980f82-05cc-46ff-bdad-2046c894d419
- Input:** Data4Prediction_V7.csv
- Output:** AutoAI-V7-Prediction

Runs (2):

Start Time	Status	Duration	Started By
Apr 29, 2021 7:44 PM	Failed	45 seconds	Gary Yeo
Apr 29, 2021 7:09 PM	Failed	45 seconds	Gary Yeo

A red circle highlights the "Status" column for the first run, which is listed as "Failed".

Deployment in AutoAI – Regression

Configure AutoAI experiment
GY_AUDUSD_Prediction1

Autosaved: 3:28:28 PM

Add data sources

Drop or browse for up to 5 tabular data files. [Learn more](#)

Browse or Select from project

MergedData_AWS.csv
Size: 0.13 MB | Columns: 8

Configure details

Create a time series forecast? [BETA](#)

Enable this option to predict future activity over a specified date/time range. Data must be structured and sequential. [Learn more](#)

Yes No

What do you want to predict?

Prediction column AUDUSD_Close

Prediction column: AUDUSD_Close CUH remaining: 20 CUH

PREDICTION TYPE: **Regression**

OPTIMIZED METRIC: RMSE

Experiment settings Run experiment

Save as

80

Pipeline leaderboard

Rank	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
1	Pipeline 4	LGBMRegressor	0.008	TFE HPO-1 FE HPO-2	00:00:40
2	Pipeline 3	LGBMRegressor	0.008	TFE HPO-1 FE	00:00:36
3	Pipeline 1	LGBMRegressor	0.009	TFE	00:00:01
4	Pipeline 2	LGBMRegressor	0.009	TFE HPO-1	00:00:13
5	Pipeline 7	DecisionTreeRegressor	0.011	TFE HPO-1 FE	00:00:23
6	Pipeline 8	DecisionTreeRegressor	0.011	TFE HPO-1 FE HPO-2	00:00:09
7	Pipeline 5	DecisionTreeRegressor	0.012	TFE	00:00:01
8	Pipeline 6	DecisionTreeRegressor	0.012	TFE HPO-1	00:00:03

Testing with AutoAI – Regression

Deployments / GY Forex Space / AUDUSD Predictions

Associated Asset Deployment GY Forex Prediction

AUDUSD Predictions Deployment Job Definition ID ee44463a-81ad-4836-84c8-51f95d3e118b Job ID 08d09037-1806-4880-84bd-21a5b2a3e202

Scheduled to run No schedule created

Environment definition 2 CPU and 8 GB RAM Edit

Input Data4Prediction_IBM.csv Edit Output AUDUSD Prediction Edit

Runs (1)

Start Time	Status	Duration	Started By
Apr 28, 2021 10:40 AM	Completed	37 seconds	Gary Yeo

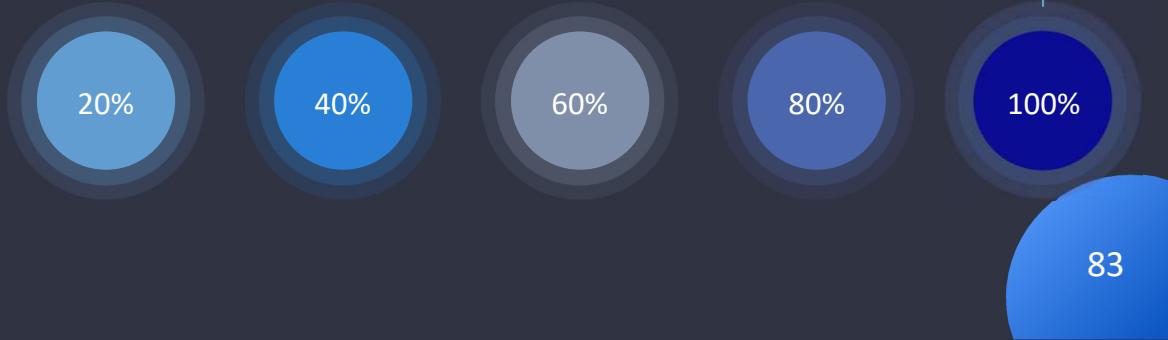
Prediction with AutoAI

A	B	C	D	E	F	G	H	I	J
1	Date	AUDUSD_Close	US_bond_2Yr	Aus_Bond_2Yr	ASX200	SP500	USDX	IRD	prediction
2	4/1/2021	0.76692	0.17	0.08	6684.2	368.79	89.87	-0.09	0.751446
3	13/1/2021	0.77461	0.14	0.08	6686.6	379.79	90.36	-0.06	0.75157
4	14/1/2021	0.77752	0.16	0.08	6715.3	378.46	90.24	-0.08	0.751814
5	15/1/2021	0.7698	0.13	0.075	6715.4	375.7	90.77	-0.055	0.756449
6	19/1/2021	0.77131	0.14	0.075	6742.6	378.65	90.5	-0.065	0.747141
7	20/1/2021	0.77544	0.13	0.075	6770.4	383.89	90.47	-0.055	0.757275
8	21/1/2021	0.77593	0.13	0.075	6823.7	384.24	90.13	-0.055	0.765155
9	22/1/2021	0.77117	0.13	0.08	6800.4	382.88	90.24	-0.05	0.762031
10	25/1/2021	0.77143	0.13	0.08	6824.7	384.39	90.39	-0.05	0.760675
11	27/1/2021	0.76478	0.12	0.085	6780.6	374.41	90.65	-0.035	0.756739
12	28/1/2021	0.76703	0.12	0.085	6649.7	377.63	90.46	-0.035	0.759604
13	29/1/2021	0.76405	0.11	0.085	6607.4	370.07	90.58	-0.025	0.75535
14	2/2/2021	0.76055	0.11	0.09	6762.6	381.55	91.2	-0.02	0.756073
15	3/2/2021	0.76359	0.13	0.09	6824.6	381.85	91.17	-0.04	0.755734
16	4/2/2021	0.75999	0.19	0.095	6765.5	386.19	91.53	-0.095	0.741505
17	16/2/2021	0.7735	0.13	0.09	6917.3	392.3	90.51	-0.04	0.755482
18	17/2/2021	0.77576	0.11	0.095	6885.2	392.39	90.95	-0.015	0.756073
19	18/2/2021	0.77678	0.11	0.105	6885.9	390.72	90.59	-0.005	0.756895
20	19/2/2021	0.78658	0.11	0.115	6793.8	390.03	90.36	0.005	0.760414
21	22/2/2021	0.79137	0.11	0.12	6780.9	387.03	90.01	0.01	0.762992

Environment Feedback



Prescriptive Analytics



Comparing Models - Versioning

Version	Python Library	Data	Results
Version 1 & 2	Sklearn 1. Linear regression 2. Multiple regression 3. Polynomial regression 4. Lasso regression 5. Ridge regression	Merged data of size 2443, after records with null value were deleted V2 had USDX removed to see if model would be able to perform better. Not significantly different.	Overfitting, unable to generalise. R square were 0.97 or more and RMSE was low
Version 3 & 4	Pycaret – auto generated models below: 1. Extra Trees Regressor 2. Random Forest Regressor 3. Light Gradient Boosting Machine 4. Decision Tree Regressor 5. Gradient Boosting Regressor 6. AdaBoost Regressor 7. K Neighbors Regressor 8. Huber Regressor 9. Orthogonal Matching Pursuit 10. Linear Regression 11. Ridge Regression 12. Bayesian Ridge 13. Passive Aggressive Regressor 14. Elastic Net 15. Lasso Regression 16. Lasso Least Angle Regression	Merged data of size 2443, after records with null value were deleted. Training & validation set separated cross validation, 10 fold, <u>randomised</u> V4 had USDX removed to see if model would be able to perform better	Overfitting, unable to generalise. R square were high at 0.99 or more and RMSE was low at 0.0057. Further tuning of model did not improve the model. Metrics appear perfect but model is useless for our prediction. USDX removal did not have significant difference on model.

Comparing Models - Versioning

Version	Python Library	Data	Results
Version 5	Pycaret – auto generated only 50% less models below suitable for time series: 1. Extra Trees Regressor 2. Decision Tree Regressor 3. AdaBoost Regressor 4. Random Forest Regressor etc	Using Pycaret auto-filling of null features with mean value which increase the sample to 3128. IRD was dropped because of negative effect on dataset. Training & validation set separated into 2011-2019 and 2020 respectively. Time series treatment – no cross validation, fold reduce to 3	R square at -2.4 and RMSE is 0.0855. Both metrics seems to suggest the model is useless. But in time series these figures are useless. but model is useless for our prediction.
Version 6	Pycaret – auto generated 9 models below suitable for time series: 1. Random Forest Regressor 2. Light Gradient Boosting Machine 3. Extra Trees Regressor 4. Gradient Boosting Regressor etc	Merged data of size 2443, after records with null value were deleted Training & validation set separated into 2011-2019 and 2020 respectively. Time series treatment – no cross validation, fold reduce to 3	Results were not within our threshold of 65%
Version 7	Pycaret – auto generated 9 models below suitable for time series: 1. Gradient Boost Regressor 2. Light Gradient Boosting Machine 3. Extra Trees Regressor 4. Random Forest Regressor etc	Using latest dataset of 3128 entries and imputed using values of prior day for the missing entry. Training & validation set separated into 2011-2019 and 2020 respectively. Time series treatment – no cross validation, fold reduce to 3	Sort by MAE (Mean Absolute Error). Results from using test data is good by using Extra Tree Regressor.

Comparing Models

Version 3

	Description	Value
0	session_id	10
1	Target	AUDUSD_Close
2	Original Data	(2443, 8)
3	Missing Values	False
4	Numeric Features	6
5	Categorical Features	0
6	Ordinal Features	False
7	High Cardinality Features	False
8	High Cardinality Method	None
9	Transformed Train Set	(2198, 25)
10	Transformed Test Set	(245, 25)
11	Shuffle Train-Test	True
12	Stratify Train-Test	False
13	Fold Generator	KFold
14	Fold Number	10

Version 7

	Description	Value
0	session_id	123
1	Target	AUDUSD_Close
2	Original Data	(2198, 10)
3	Missing Values	False
4	Numeric Features	8
5	Categorical Features	1
6	Ordinal Features	False
7	High Cardinality Features	False
8	High Cardinality Method	None
9	Transformed Train Set	(2198, 20)
10	Transformed Test Set	(245, 20)
11	Shuffle Train-Test	True
12	Stratify Train-Test	False
13	Fold Generator	TimeSeriesSplit
14	Fold Number	3



Comparing Models

Version 3

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
et	Extra Trees Regressor	0.0048	0.0000	0.0067	0.9975	0.0036	0.0058	0.2320
rf	Random Forest Regressor	0.0057	0.0001	0.0088	0.9956	0.0047	0.0068	0.2550
lightgbm	Light Gradient Boosting Machine	0.0071	0.0001	0.0103	0.9938	0.0055	0.0083	0.1500
dt	Decision Tree Regressor	0.0073	0.0001	0.0114	0.9926	0.0061	0.0087	0.0100
gbr	Gradient Boosting Regressor	0.0104	0.0002	0.0142	0.9887	0.0075	0.0123	0.1120
ada	AdaBoost Regressor	0.0199	0.0006	0.0244	0.9665	0.0131	0.0238	0.0670
knn	K Neighbors Regressor	0.0216	0.0009	0.0306	0.9472	0.0163	0.0253	0.0110
br	Bayesian Ridge	0.0483	0.0047	0.0685	0.7370	0.0349	0.0543	0.0090
ridge	Ridge Regression	0.0485	0.0048	0.0689	0.7335	0.0351	0.0545	0.0070
lr	Linear Regression	0.0485	0.0048	0.0690	0.7331	0.0351	0.0545	0.6590
lar	Least Angle Regression	0.0485	0.0048	0.0690	0.7331	0.0351	0.0545	0.0080
huber	Huber Regressor	0.0486	0.0051	0.0704	0.7197	0.0357	0.0547	0.4240
omp	Orthogonal Matching Pursuit	0.0500	0.0058	0.0756	0.6793	0.0378	0.0556	0.0070
en	Elastic Net	0.0843	0.0106	0.1030	0.4076	0.0545	0.0964	0.0080
lasso	Lasso Regression	0.1152	0.0187	0.1366	-0.0418	0.0730	0.1330	0.0080
llar	Lasso Least Angle Regression	0.1152	0.0187	0.1366	-0.0418	0.0730	0.1330	0.0070

	Date	AUDUSD_Close	US_bond_2Yr	Aus_Bond_2Yr	ASX200	SP500	IRD	Label
0	2021-02-03	0.78238	0.13	0.090	6824.6	381.85	-0.040	0.698492
1	2021-03-03	0.77525	0.14	0.110	6818.0	381.42	-0.030	0.697427
2	2021-03-15	0.77475	0.14	0.085	6773.0	396.41	-0.055	0.697347
3	2021-03-16	0.77457	0.15	0.080	6827.1	395.91	-0.070	0.697631
4	2021-03-17	0.78140	0.13	0.080	6795.2	397.26	-0.050	0.697347
5	2021-03-18	0.77520	0.16	0.085	6745.9	391.48	-0.075	0.697631
6	2021-03-19	0.77369	0.16	0.085	6708.2	389.48	-0.075	0.697631
7	2021-03-22	0.77443	0.15	0.080	6752.5	392.59	-0.070	0.697631
8	2021-03-23	0.76062	0.15	0.080	6745.4	389.50	-0.070	0.697631
9	2021-03-24	0.75903	0.14	0.075	6778.8	387.52	-0.065	0.697631
10	2021-03-25	0.75921	0.14	0.070	6790.6	389.70	-0.070	0.697631
11	2021-03-26	0.76371	0.14	0.070	6824.2	395.98	-0.070	0.697631

Predictions Using New Data

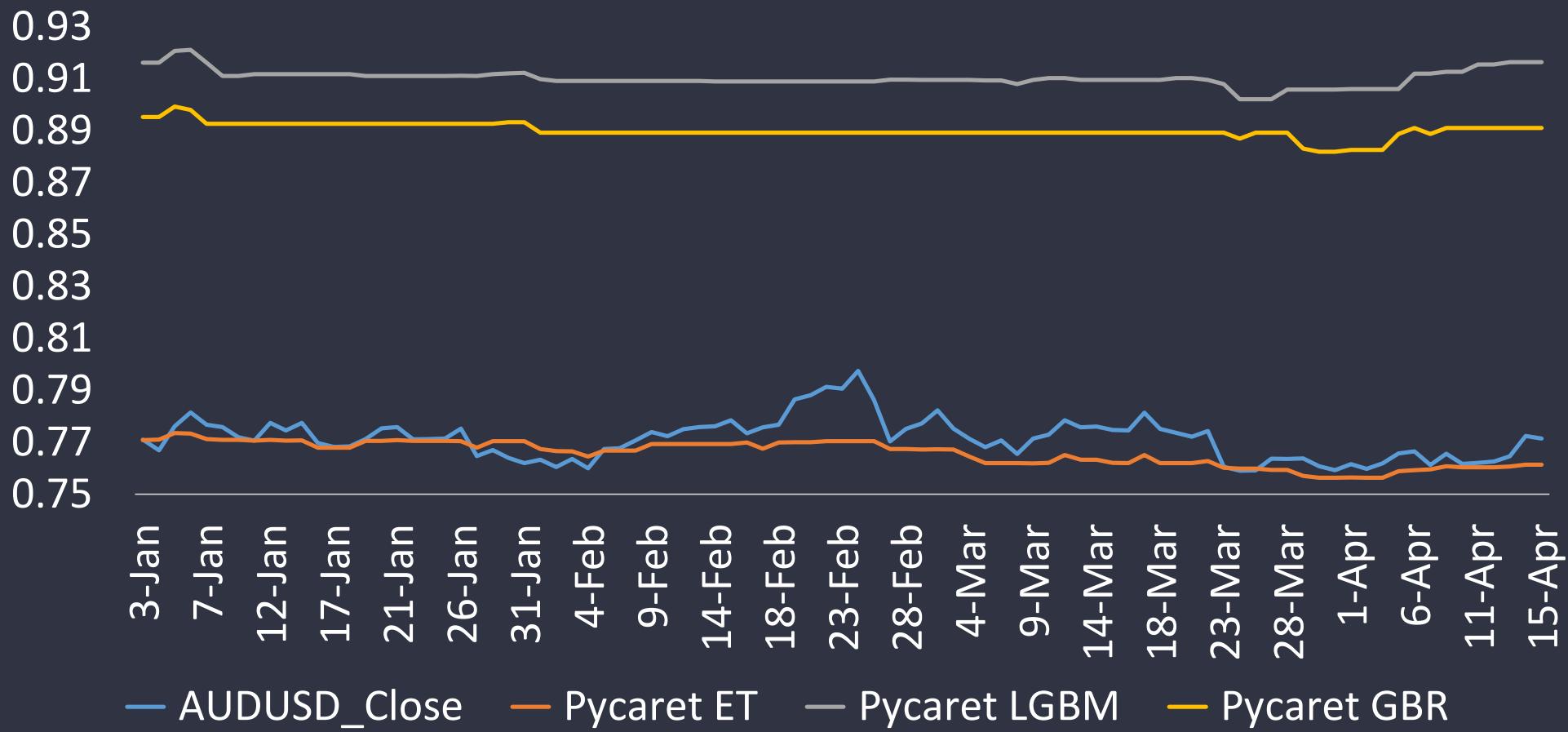
Version 7

Model		MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
gbr	Gradient Boosting Regressor	0.0722	0.0094	0.0813	-2.0559	0.0439	0.0905	0.2367
lightgbm	Light Gradient Boosting Machine	0.0722	0.0094	0.0812	-2.0213	0.0438	0.0907	0.4600
et	Extra Trees Regressor	0.0733	0.0096	0.0832	-2.2385	0.0449	0.0920	0.3100
rf	Random Forest Regressor	0.0745	0.0099	0.0843	-2.2961	0.0455	0.0939	1.6967
knn	K Neighbors Regressor	0.0759	0.0102	0.0881	-2.9383	0.0476	0.0953	2.1900
dt	Decision Tree Regressor	0.0771	0.0100	0.0873	-2.7340	0.0474	0.0979	1.6200
ada	AdaBoost Regressor	0.0784	0.0106	0.0879	-2.9375	0.0473	0.0984	0.1300
par	Passive Aggressive Regressor	0.1388	0.0235	0.1486	-12.7557	0.0822	0.1779	1.8400
llar	Lasso Least Angle Regression	0.1792	0.0357	0.1861	-35.9590	0.1001	0.2355	1.8367

Date	AUDUSD_Close	Pycaret ET	Pycaret LGBM	Pycaret GBR
3/1/2021	0.77109	0.770871	0.916226	0.895288
4/1/2021	0.76692	0.771059	0.916226	0.895288
5/1/2021	0.77618	0.773591	0.920757	0.899299
6/1/2021	0.78152	0.773383	0.921067	0.898017
7/1/2021	0.77682	0.771237	0.916226	0.892628
8/1/2021	0.77591	0.771018	0.911076	0.892628
10/1/2021	0.77202	0.771018	0.911076	0.892628
11/1/2021	0.77068	0.77071	0.911771	0.892628
12/1/2021	0.77752	0.771018	0.911771	0.892628
13/1/2021	0.77461	0.770646	0.911771	0.892628
14/1/2021	0.77752	0.770806	0.911771	0.892628

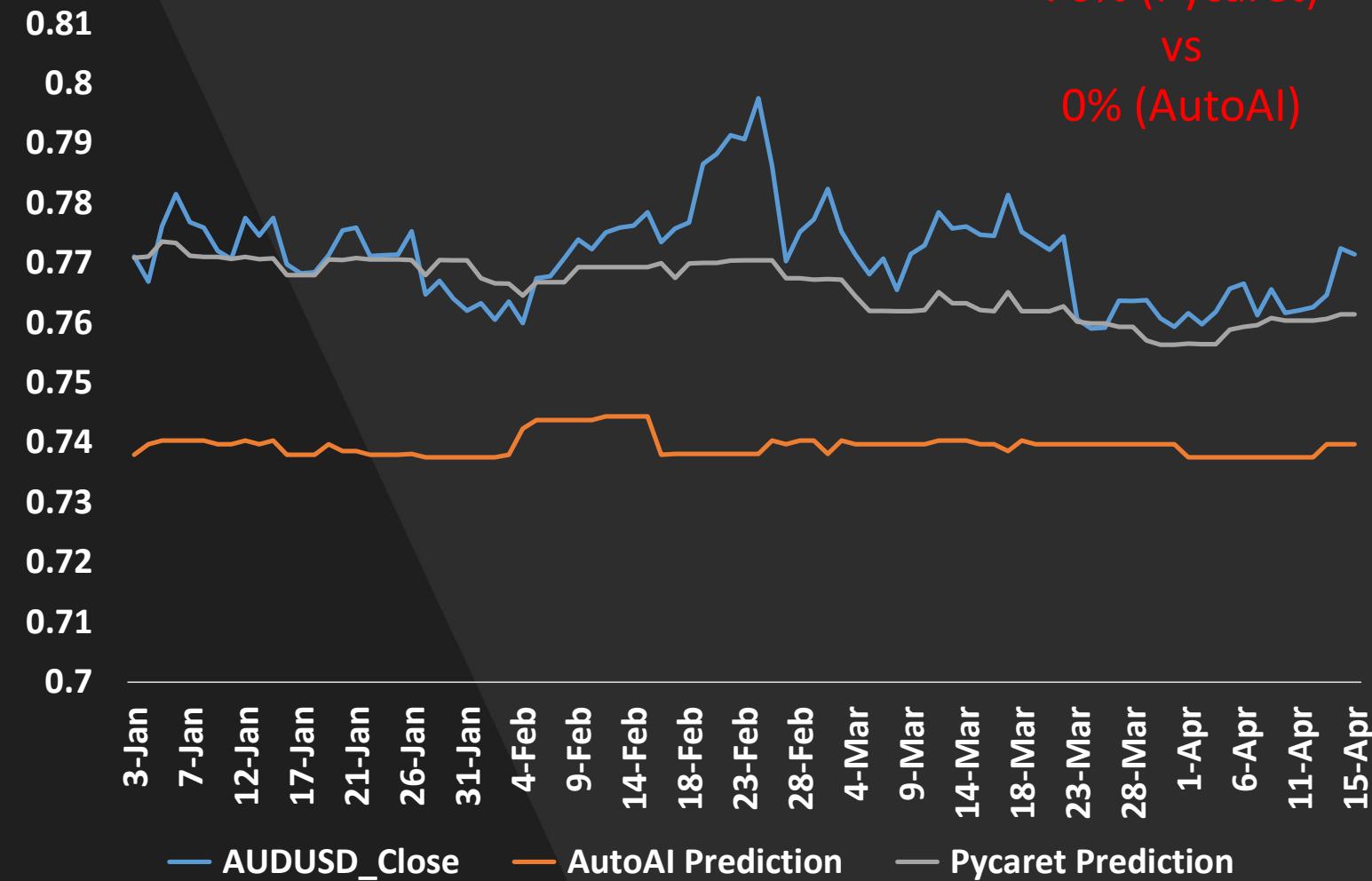
Predictions Using New Data

Q1 2021 Predictions



Pycaret vs AutoAI

Accuracy of
76% (Pycaret)
vs
0% (AutoAI)



A	B	C	D	
Date	AUDUSD_Close	AutoAI Predicti	Pycaret Predicti	
2	3/1/2021	0.77109	0.738000313	0.770871
3	4/1/2021	0.76692	0.739756381	0.771059
4	5/1/2021	0.77618	0.740367679	0.773591
5	6/1/2021	0.78152	0.740367679	0.773383
6	7/1/2021	0.77682	0.740367679	0.771237
7	8/1/2021	0.77591	0.740367679	0.771018
8	10/1/2021	0.77202	0.739756381	0.771018
9	11/1/2021	0.77068	0.739756381	0.77071
10	12/1/2021	0.77752	0.740367679	0.771018
11	13/1/2021	0.77461	0.739756381	0.770646
12	14/1/2021	0.77752	0.740367679	0.770806
13	15/1/2021	0.7698	0.738000313	0.768004
14	17/1/2021	0.76828	0.738000313	0.768004
15	18/1/2021	0.76846	0.738000313	0.767984
16	19/1/2021	0.77131	0.739756381	0.770598
17	20/1/2021	0.77544	0.738611611	0.770524
18	21/1/2021	0.77593	0.738611611	0.770831
19	22/1/2021	0.77117	0.738000313	0.770619
20	24/1/2021	0.7713	0.738000313	0.770619
21	25/1/2021	0.77143	0.738000313	0.770592
22	26/1/2021	0.77529	0.738136374	0.770513
23	27/1/2021	0.76478	0.737525077	0.768004
24	28/1/2021	0.76703	0.737525077	0.770512
25	29/1/2021	0.76405	0.737525077	0.770458
26	31/1/2021	0.76207	0.737525077	0.770458
27	1/2/2021	0.76331	0.737525077	0.767485
28	2/2/2021	0.76055	0.737525077	0.7666
29	3/2/2021	0.76359	0.738000313	0.766564
30	4/2/2021	0.75999	0.74237318	0.764563
31	5/2/2021	0.76748	0.743773338	0.766818
32	7/2/2021	0.7678	0.743773338	0.766818
33	8/2/2021	0.77079	0.743773338	0.766818
34	9/2/2021	0.77395	0.743773338	0.769343
35	10/2/2021	0.77233	0.743773338	0.769347
36	11/2/2021	0.77513	0.744384635	0.769343
37	12/2/2021	0.77592	0.744384635	0.769343
38	14/2/2021	0.77624	0.744384635	0.769343
39	15/2/2021	0.7785	0.744384635	0.769343
40	16/2/2021	0.7735	0.738000313	0.769999

In order to succeed, you first have to be willing to experience failure."

— Yvan Byeajee, The essence of trading psychology in one skill

THANK YOU

For Your Attention & Appreciation

