BUSINESS KNOWLEDGE FOR DATA SCIENCE

-Presented by Anurag Soin

WHAT TO EXPECT

1. What we expect vs what it is

2. How to work under such a scenario: Framework

3. Identify the focus Area

    1. Growth Areas: New Opportunities

    2. Focus areas: biggest margin/profit

    3. Issues areas: improvements/solutions

GROUP FORMATION

WHAT IS YOUR VIEW AROUND DATA FLOW IN A BIG SOPHISTICATED ORGANISATION & WHERE DATA SCIENTIST FIT IN

DATA ISSUES

WHAT WE EXPECT

Data Generation

Data Flow

Data Collection

Process Improvement

Databases

Reporting
- what *did* happened *yesterday*?

Data / BI Analyst ?
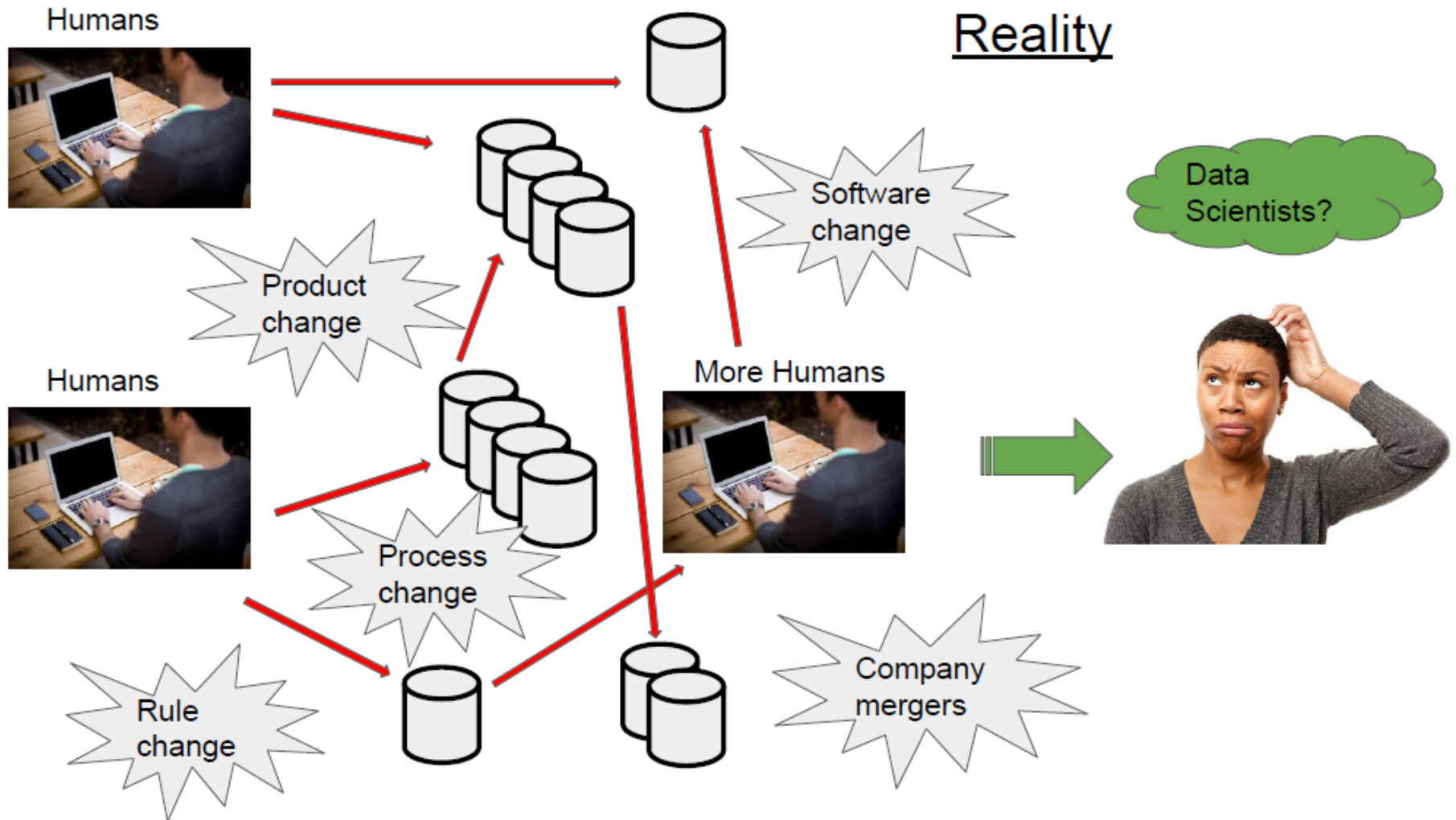
Predictive Modelling
- what *might* happen *tomorrow*?

Process Optimisation
- what *should* happen *tomorrow*?

Database Administrator ?

Data Scientist ?

REALITY

Humans

Humans

Reality

Product change

Process change

Rule change

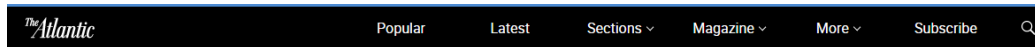Software change

More Humans

Company mergers

Data Scientists?

6

PEOPLE HAVE A WAY TO REACT: AI & ML

Perception

High-Speed Trading Isn't About Efficiency—It's About Cheating

Stockmarkets rigged by high-speed traders, claims Moneyball author Michael Lewis
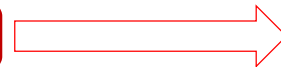
HUFFPOST AUSTRALIA

Now lets look at Chess

One Step at a time

- Rules well defined.
- Players are allowed to research opponents playing tactics data
- Players are allowed to play as many practice games they want
- Players are allowed to invest in the upskilling and any associated technology
- Who Wins: The golden rule of capitalism the Skills, Efforts and out of box thinking always get rewarded

WE LIVE IN A DYNAMIC WORLD: DATA SCIENCE IS MORE ABOUT THOUGHT PROCESS THAN BUILDING MODELS

- Modelling is only as good as the data
- Model building is now becoming point and click $\longrightarrow$ **Thought Process**
- 'Correct' data preparation will never be this
- Rubbish in - Rubbish out
- Caveat Emptor
  - *let the buyer beware*
  - *don't trust any data*

Data Sanity Checking - Identifying Systematic Data Issues

Predictive modelling assumes the future will be like the past - we first have to make sure the past is like the past

- Knowing your data is the most important thing
- 'expert opinion' is only opinion
- The data contains the questions
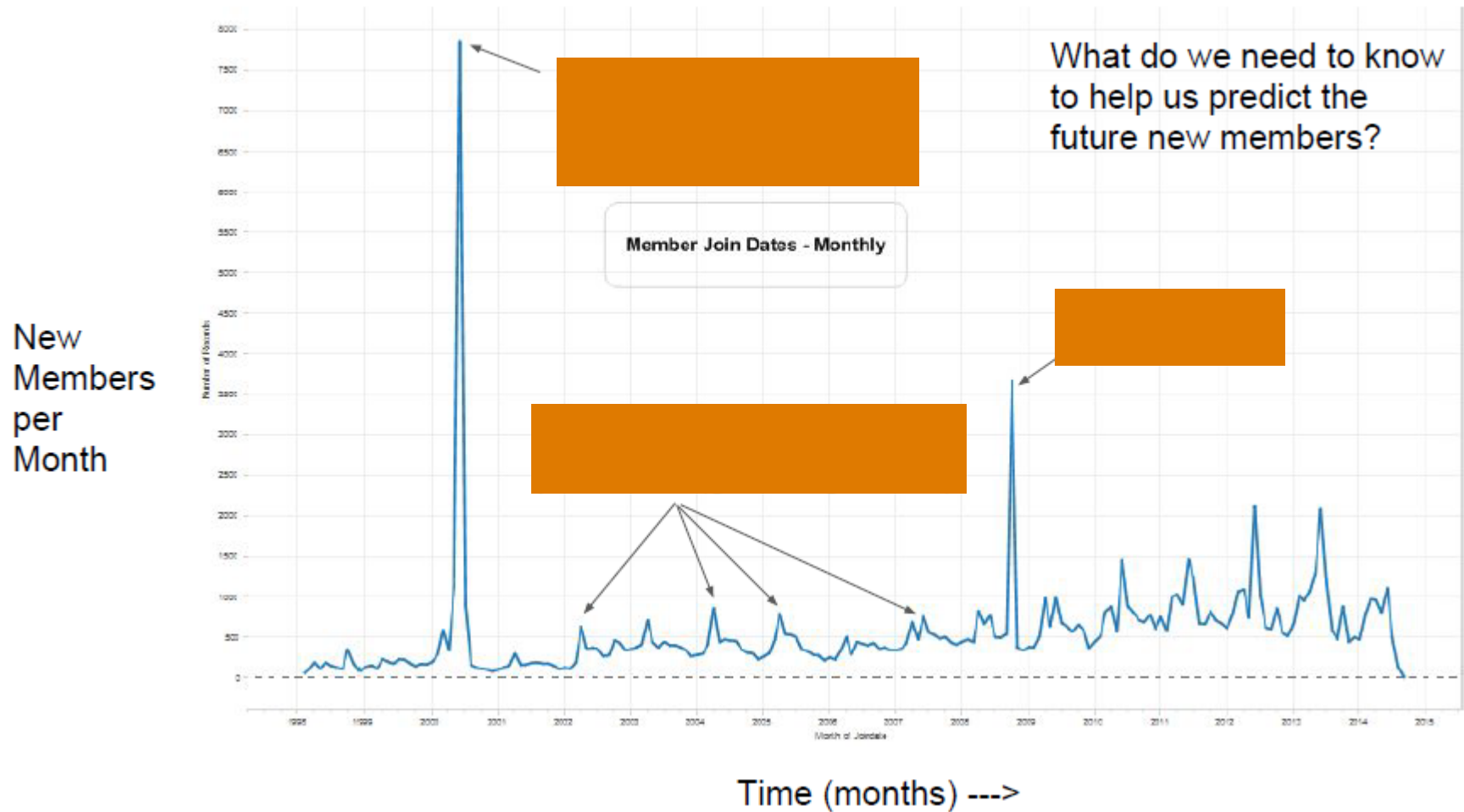 - *the 'experts' may have the answers*

Don't forget about IBM watson storey

DATA ISSUES
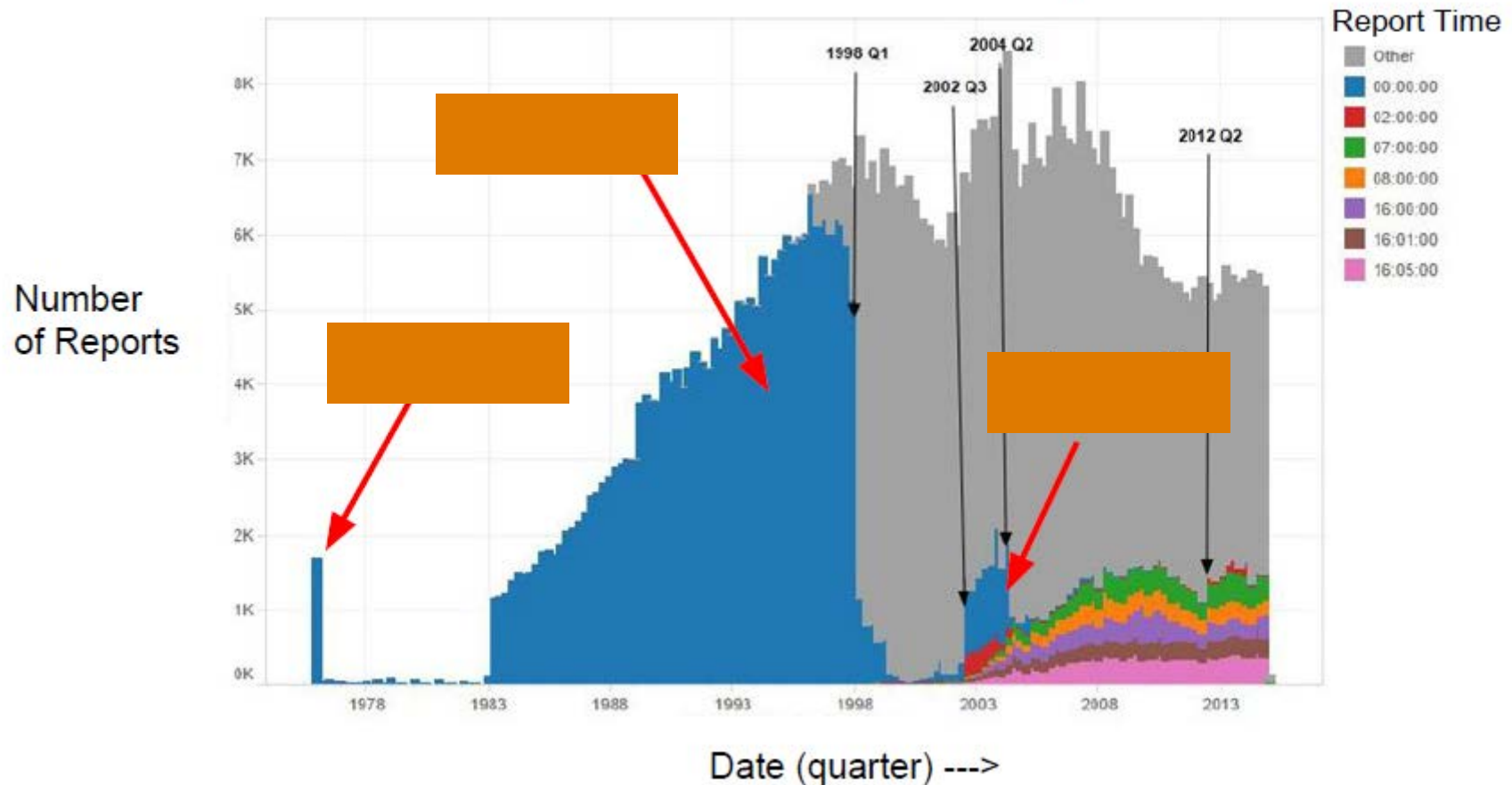
PROBLEMS OF DYNAMIC WORLD

Always go back and ask why

PROBLEMS OF DYNAMIC WORLD

If the data has more than one dimension always plot

Company Financial Statement Dates

PEOPLE ISSUES

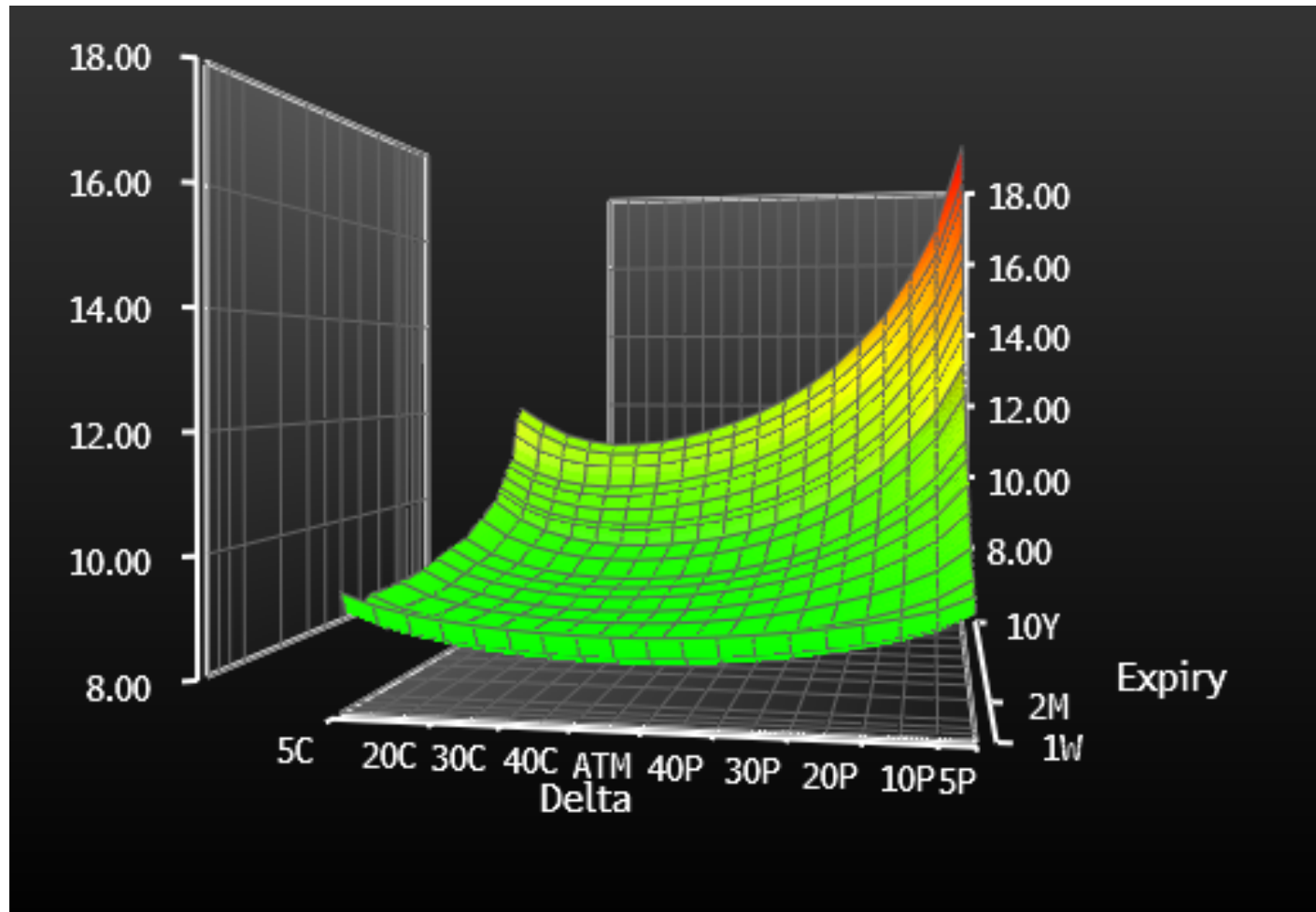# AUDUSD: EXPERT OPINION IS JUST AN OPINION

AUDUSD: BUILD SYSTEMS TO BE SUCCESSFUL

AUDUSD	↑ .8019	+.0016				BGN · .8018 / .8019		BGN	
At 15:34	Op .8003	Hi .8043	Lo .7996			Close .8003		Value 08/03/17	

AUDUSD Curncy   Add ▾   Review   Actions ▾   Backtesting & Optimization: Multi Analysis

08/01/2016 — 08/01/2017   ☑ Today   Daily ▾

Table | Chart | Scatter Plot | Graphics Filter

Worksheet Title BTST   « ¶ Strategies ⋎ Columns ▶ Playlist

Strategy				Current Position		Trades			Profit (Loss)				Statistics			
				P&L	Age	Long	Short	Total	Total	%Total	Long	Short	Avg Duration	Profit Factor	Sharpe	Sort Rat
Maximum				6.83M	261	68	68	136	9.41M	9.41	7.89M	1.52M	261.00	1.86	1.27	1.47
Average				398.73k	46	15	15	30.83	78.55k	0.08	3.25M	-3.17M	32.25	0.95	0.01	0.01
Median				414.33k	39	10	10	20	-899.69k	-0.90	2.79M	-3.38M	13.05	0.90	-0.13	-0.1
Minimum				-7.58M	2	1	0	1	-7.28M	-7.28	-468.92	-6.81M	2.91	0.00	-0.99	-1.1
31) Buy & Hold			▲	6.40M	261	1	0	1	6.40M	6.40	6.40M	0.00	261.00	0.00	0.88	0.97
32) Bollinger Bands (Bo	✎	⊗	▼	-6.42M	39	3	3	6	-3.09M	-3.09	1.97M	-5.05M	40.83	0.66	-0.42	-0.4
33) Cmdty Channel Inde	✎	⊗	▼	-7.37M	36	8	9	17	9.15M	9.15	7.70M	1.45M	15.76	1.86	1.25	1.42
34) DMI	✎	⊗	▲	6.81M	41	15	15	30	2.87M	2.87	4.03M	-1.16M	9.17	1.22	0.40	0.48
35) MACD	✎	⊗	▲	3.38M	13	13	12	25	-5.55M	-5.55	-246.28	-5.31M	11.36	0.70	-0.74	-0.9
36) RSI	✎	⊗	▼	-4.86M	126	1	1	2	-909.21k	-0.91	3.95M	-4.86M	91.00	0.81	-0.16	-0.1
37) Stochastics (TAS)	✎	⊗	▼	-7.08M	29	7	8	15	8.22M	8.22	7.86M	358.83k	18.33	1.79	1.13	1.35
38) William's %R (Wm)	✎	⊗	▼	-7.58M	37	8	9	17	9.41M	9.41	7.89M	1.52M	15.76	1.71	1.27	1.47
39) Parabolic (PTPS)	✎	⊗	▲	3.50M	13	13	12	25	-2.21M	-2.21	1.55M	-3.75M	11.24	0.87	-0.30	-0.3
40) Simple MA (SMAvg)	✎	⊗	▲	4.97M	17	16	16	32	-4.12M	-4.12	427.10k	-4.55M	8.63	0.77	-0.56	-0.6
41) Exponential MA (EMA	✎	⊗	▲	6.57M	41	12	12	24	-866.37k	-0.87	2.47M	-3.34M	11.04	0.93	-0.11	-0.1
42) Weighted MA (WMAv	✎	⊗	▲	6.67M	41	10	10	20	723.79k	0.72	3.40M	-2.68M	13.05	1.05	0.09	0.10
43) Variable MA (VMAvg	✎	⊗	▲	6.58M	41	16	16	32	-625.54k	-0.63	2.67M	-3.29M	8.53	0.96	-0.08	-0.1
44) Triangular MA (TMA	✎	⊗	▲	6.55M	41	9	9	18	-1.07M	-1.07	2.79M	-3.86M	14.28	0.92	-0.16	-0.1
45) Accum/Distrib Osc	✎	⊗	▲	414.33k	2	68	68	136	670.80k	0.67	3.81M	-3.14M	2.91	1.02	Ac 0.10	0.12

ANZ

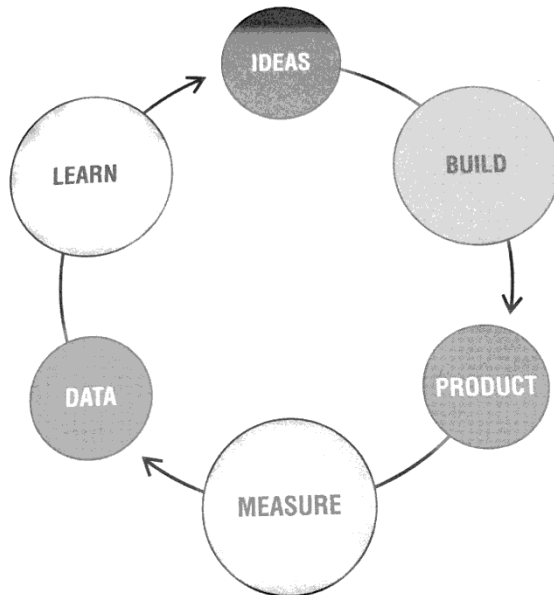# AUDUSD: ALWAYS TAKE SOME OTHER VIEW INTO CONSIDERATION

DAILY ISSUES

1. Problem statement is not well defined

2. The delivery is beyond your circle of influence

3. Time pressure without understanding the complexity

4. How you explain the black box, if people cant understand it they will

    not use it

ANZ

How to work under such scenario

THE ONLY TIME I HAVE DONE IT RIGHT

"Planning is a tool that only works in the presence of a long and stable history"

BUILD-MEASURE-LEARN FEEDBACK LOOP

Minimize *TOTAL* time through the loop

1. Genchi gembutsu....first hand info

2. *Get out of the building*

3. *Goal is not to get definitive answer*

4. *Paralysis by Analysis vs Just do it*

How to work under such scenario

THE ONLY TIME I HAVE DONE IT RIGHT

- Identify the focus Area:

    - Growth Areas: New Opportunities--> **New Revenue Streams (ANZ)**

    - Focus areas: biggest margin/profit--> **Hotel Spending (NRMA)**

    - Issues areas: improvements/solutions--> **Upstream Business Risk (BP)**