Capstone Time Series Project

“Time Series / Forecasts:

From the basic solution to the complex

– daily and monthly

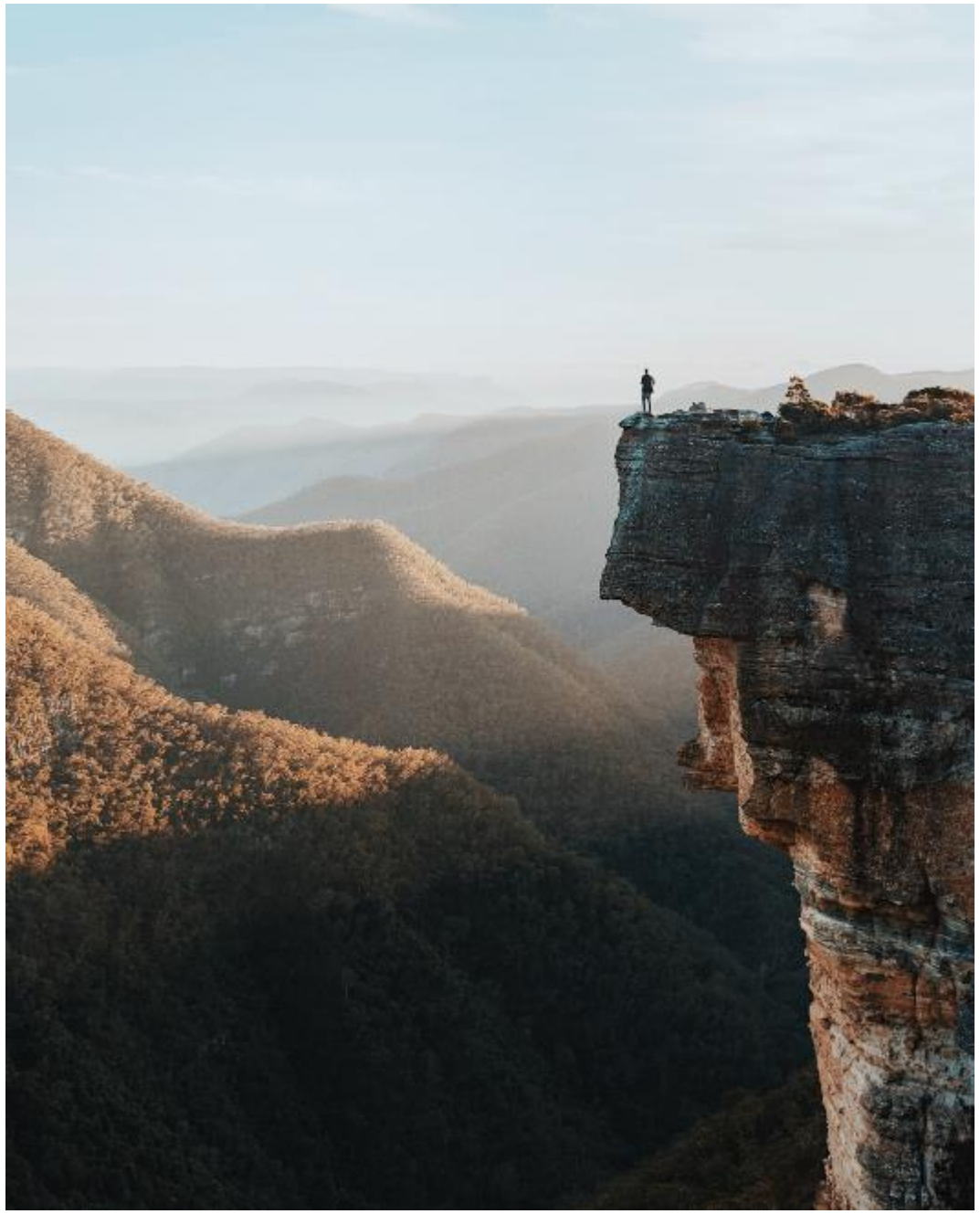
- by store and by product.”

by Alex Dance

Agenda

1. Exploratory data analysis
2. The business problem
3. The approach
4. The results
5. Next steps

EXPLORATORY DATA ANALYSIS



Time Series Problem:

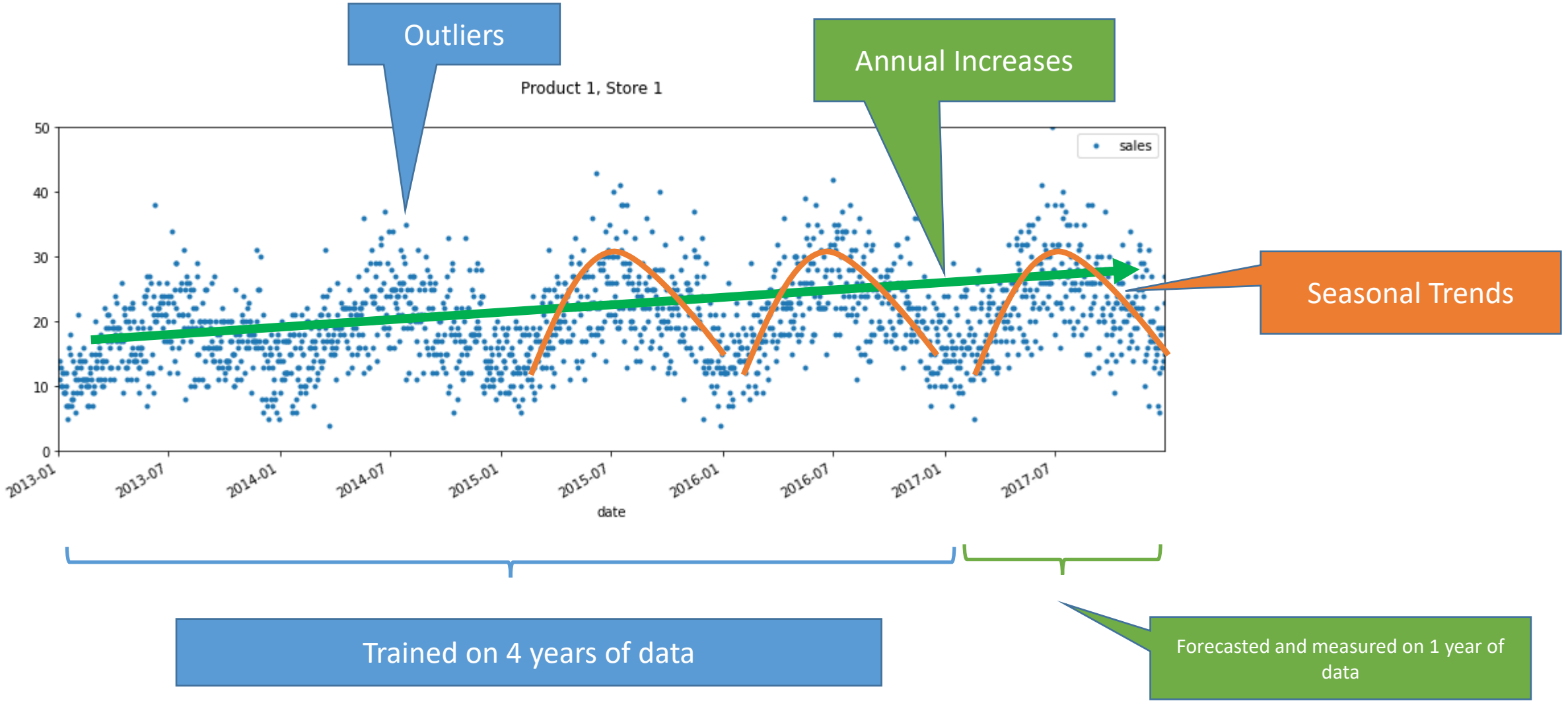
Have a data set of 5 years of retail sales (917K rows)

- by day
- by product (50 products)
- by store (10 stores)

totaling over 10M sales p/a  **at $50 a sale = $500M sales per annum**

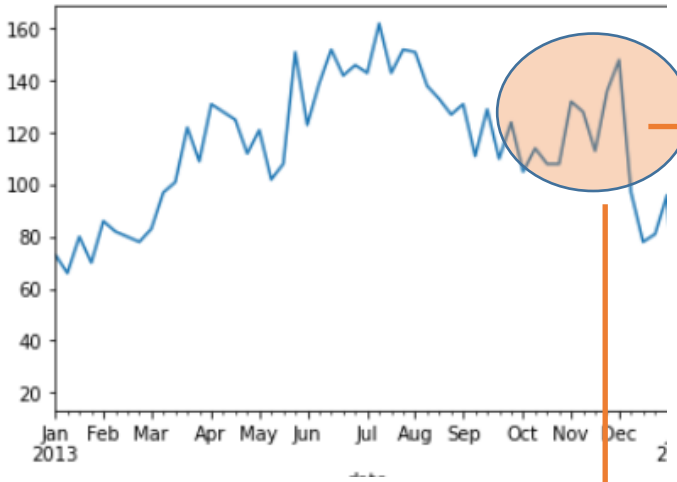
Goal: Do a **better** job of forecasting future sales -> Improve Profitability

A quick look at the data highlights annual increases and seasonal trends

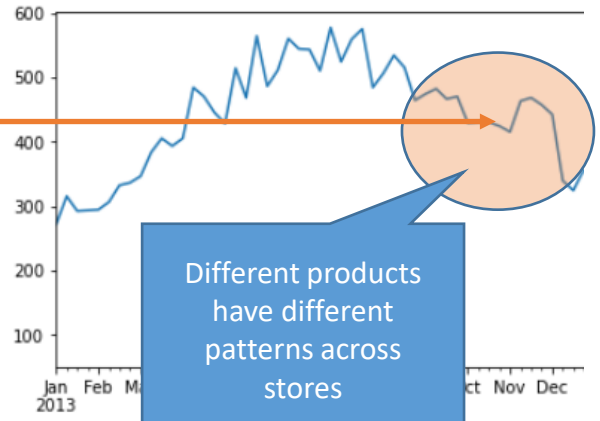


Adding to the complexity there is variability between stores and items

Item 1 Store 1

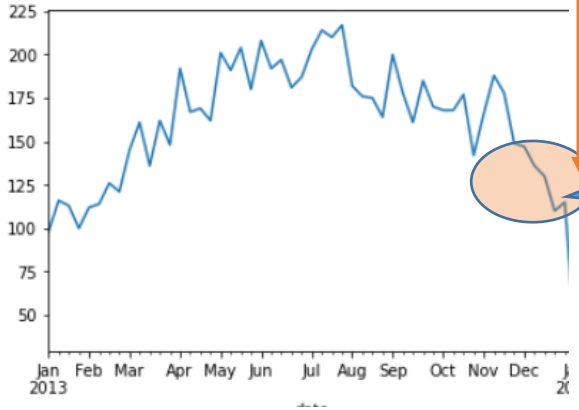


Item 1 Store 2



Different products have different patterns across stores

Item 2 Store 1



Stores don't have the same patterns for all products

Hardest to do/ best forecast
IS

by store
by product
By day

10 stores + 50 products = 500 forecasts by day over a year = 182,500 values

BUSINESS PROBLEM

A better forecast will lead to improved profitability PLUS a better customer experience

Issue	Details	Assumptions	Annual Benefit
Customer Need	Customers can't find products they want	7% increase when would have been a lost sale (4% of the time)	$1.4M
Use of space	• Better use of store floor space • Better use of warehouse space	3% less spend on rent	$1.5M
Organisation Level	Better demand Planning	1% reduced staff costs	$750K
Discounting	Less stock clearance sales	15% less discounting	$1.25M
Not Sold	Less Waste	25% less throw away stock	$2.5M
			$7.4M

Based on a previous 10% EBITDA this will improve EBITDA by 14.85%

TOTAL BENEFIT is 1.5% of total sales ($500M pa)

THE APPROACH

Expanded the date, looked back and added holidays

Expanded Date

From: 14 Jan 2013
To: Extra Features

- Day of week
- Day of Year
- Month
- Day of Month
- Year

Looked at previous days

Forecast 1 January by looking at the trend over the last 7 days

Added external data (holidays)

Ran multiple AI models to solve this time series problem

Feature Engineering + **Adjusting Data** + **Artificial Intelligence Algorithms** + **Amazon/ AWS Forecast + Sagemaker**

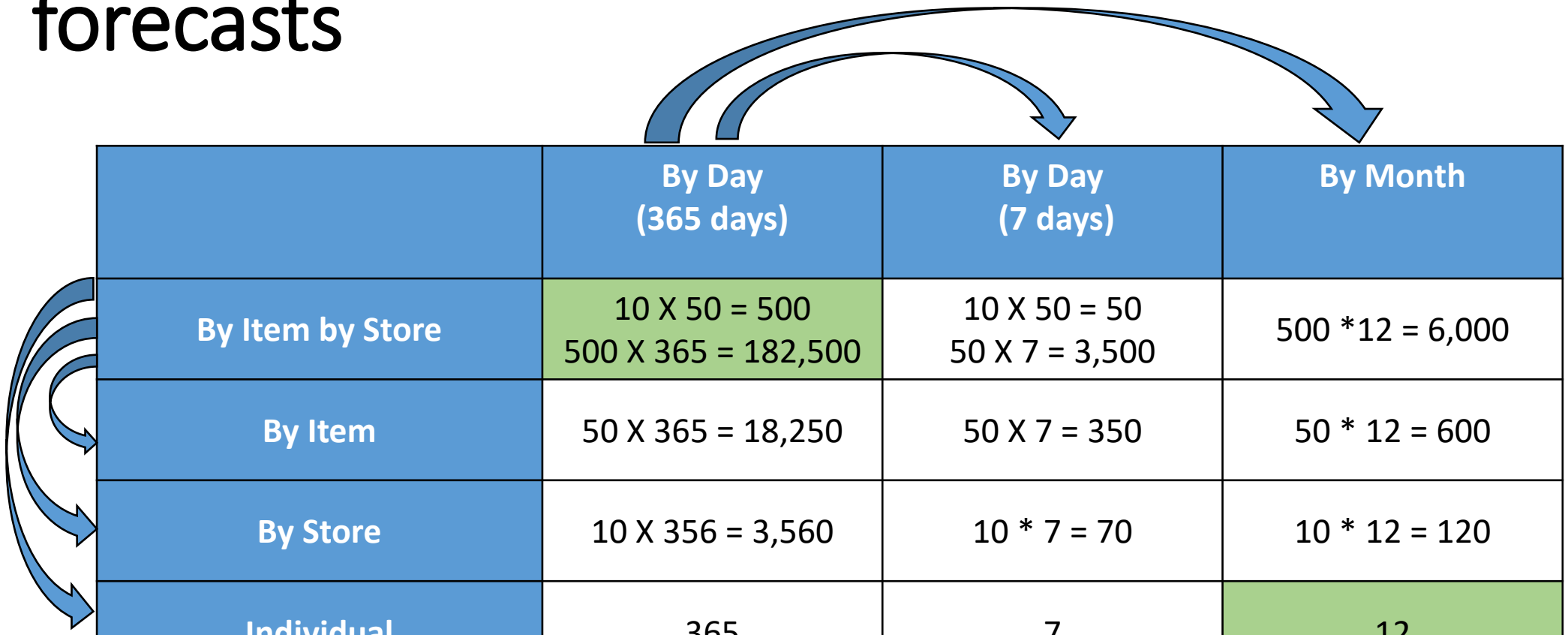
Simple → Complex

Rolling Averages	Weighted Average across all the data	ARIMA	XG Boost	Deep Learning
Weighted Moving Averages		Seasonal ARIMA	Cat Boost	
Smoothed Moving Averages		AR	Prophet	

For each NEED choose the

Best Forecast

Able to meet different needs / periods across forecasts

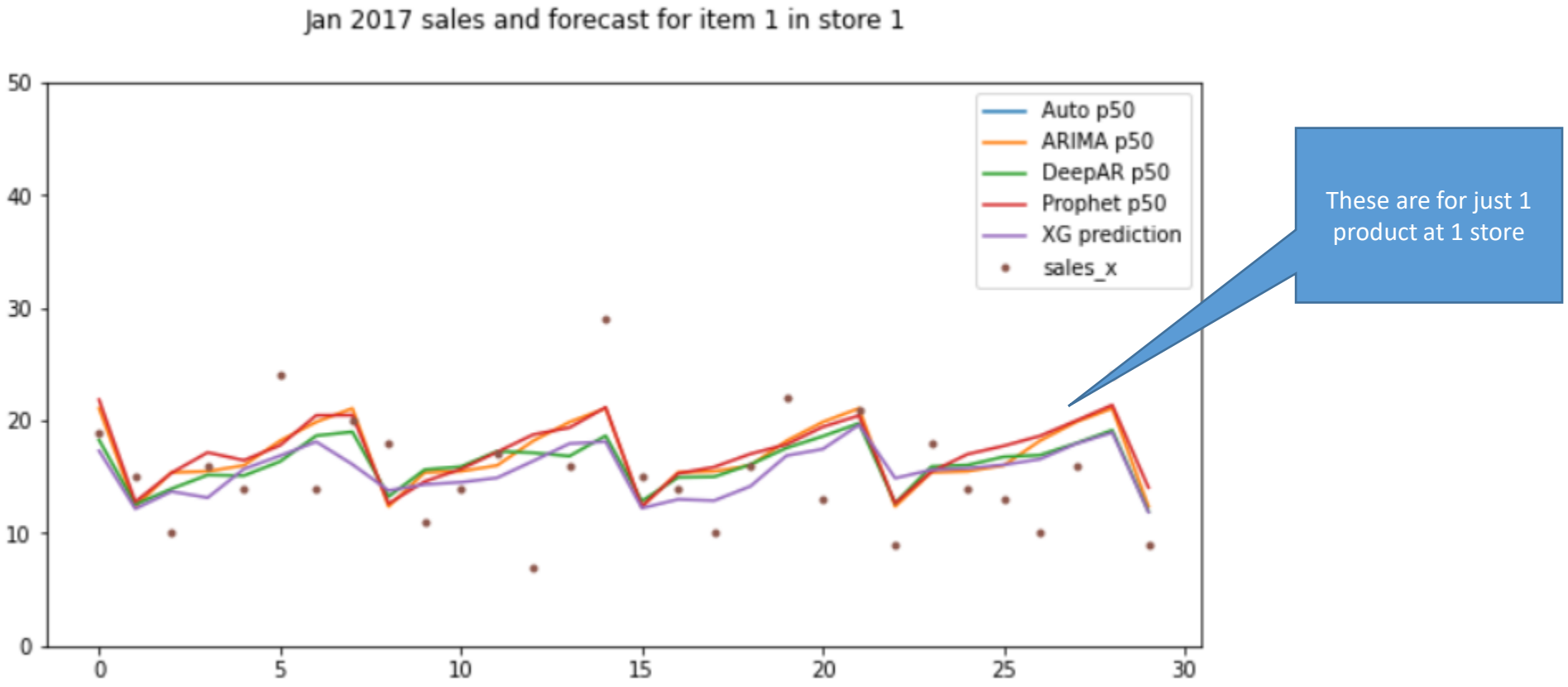


	By Day (365 days)	By Day (7 days)	By Month
By Item by Store	10 X 50 = 500 500 X 365 = 182,500	10 X 50 = 50 50 X 7 = 3,500	500 *12 = 6,000
By Item	50 X 365 = 18,250	50 X 7 = 350	50 * 12 = 600
By Store	10 X 356 = 3,560	10 * 7 = 70	10 * 12 = 120
Individual	365	7	12

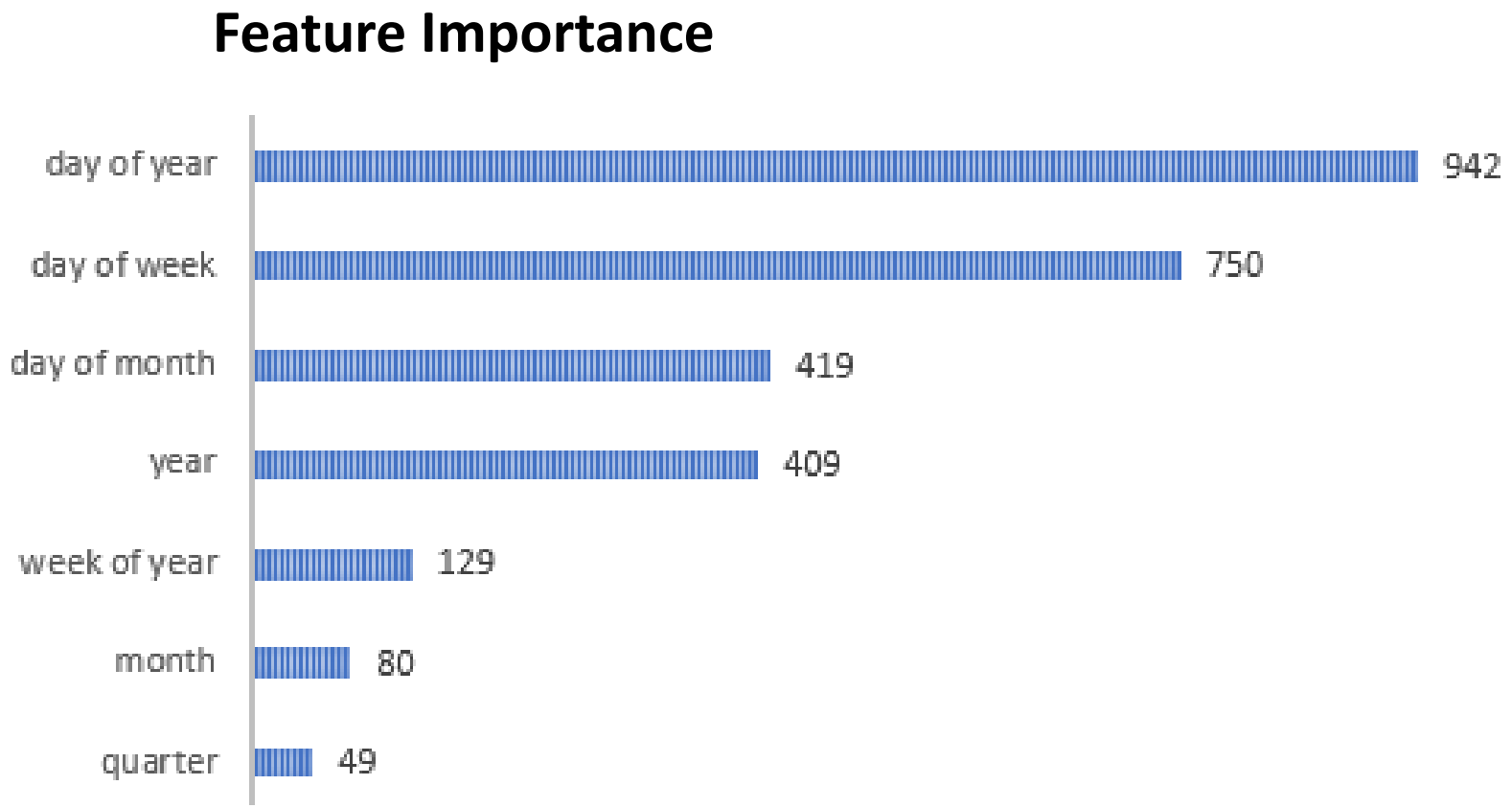
Plus a GRAND TOTAL Forecast

THE RESULTS

Compared multiple models to see variations – such as weekly variations



Know the importance of different features



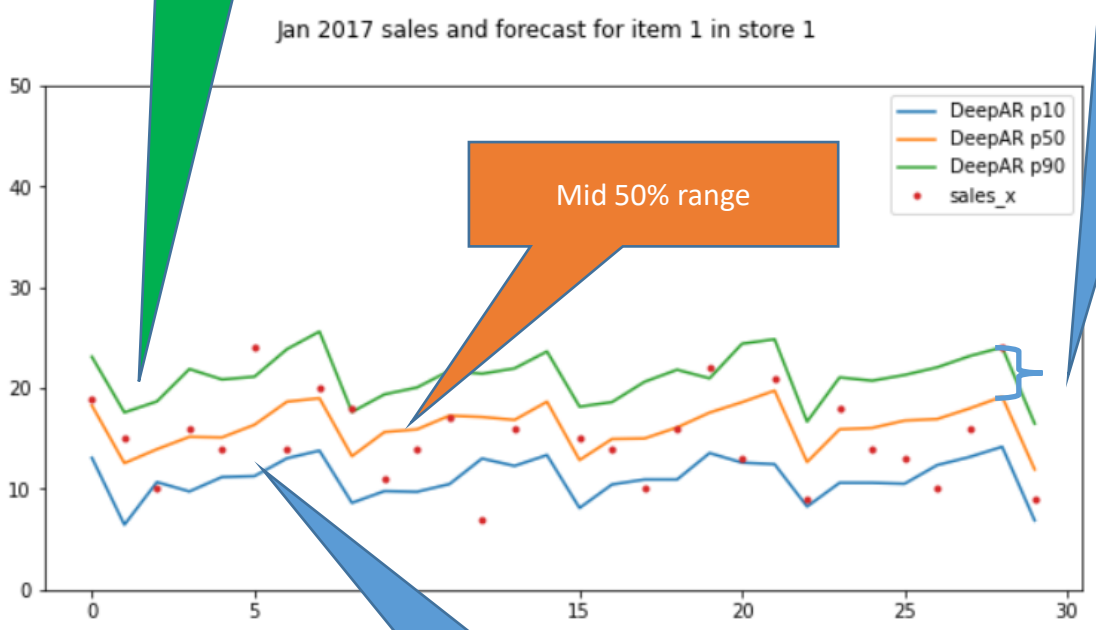
The forecasts confidence levels are still as good in December as they were for January

Top 90% range

Mid 50% range

Bottom 10% range

Jan 2017

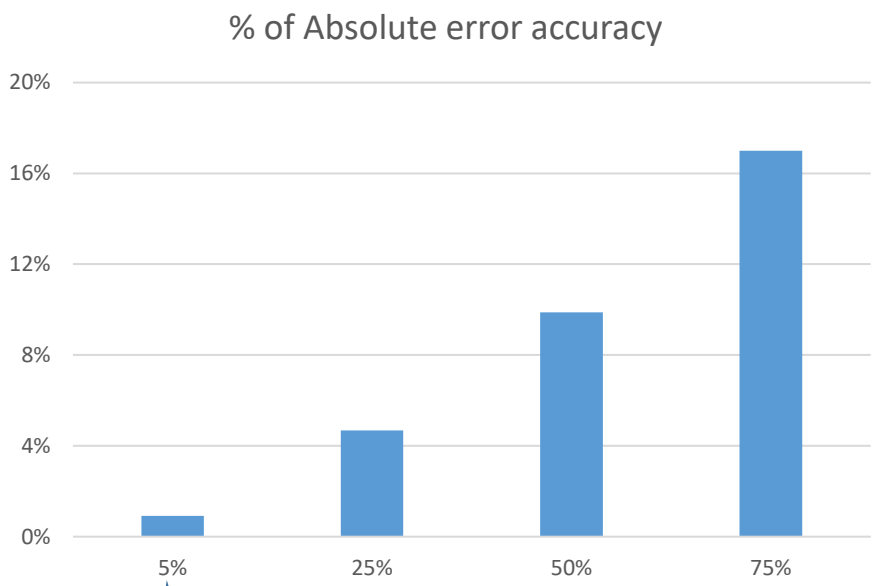


Confidence intervals have not expanded out

Nov and Dec 2017

Still picking up little dips and blips accurately

For each model there are multiple metrics to potentially review



5% of forecasts are within 0.92% of actual sales

75% of forecasts are within 17% of actual sales

XG Boost for product / store over 365 days
Root Mean Squared Error (RMSE)
6.31

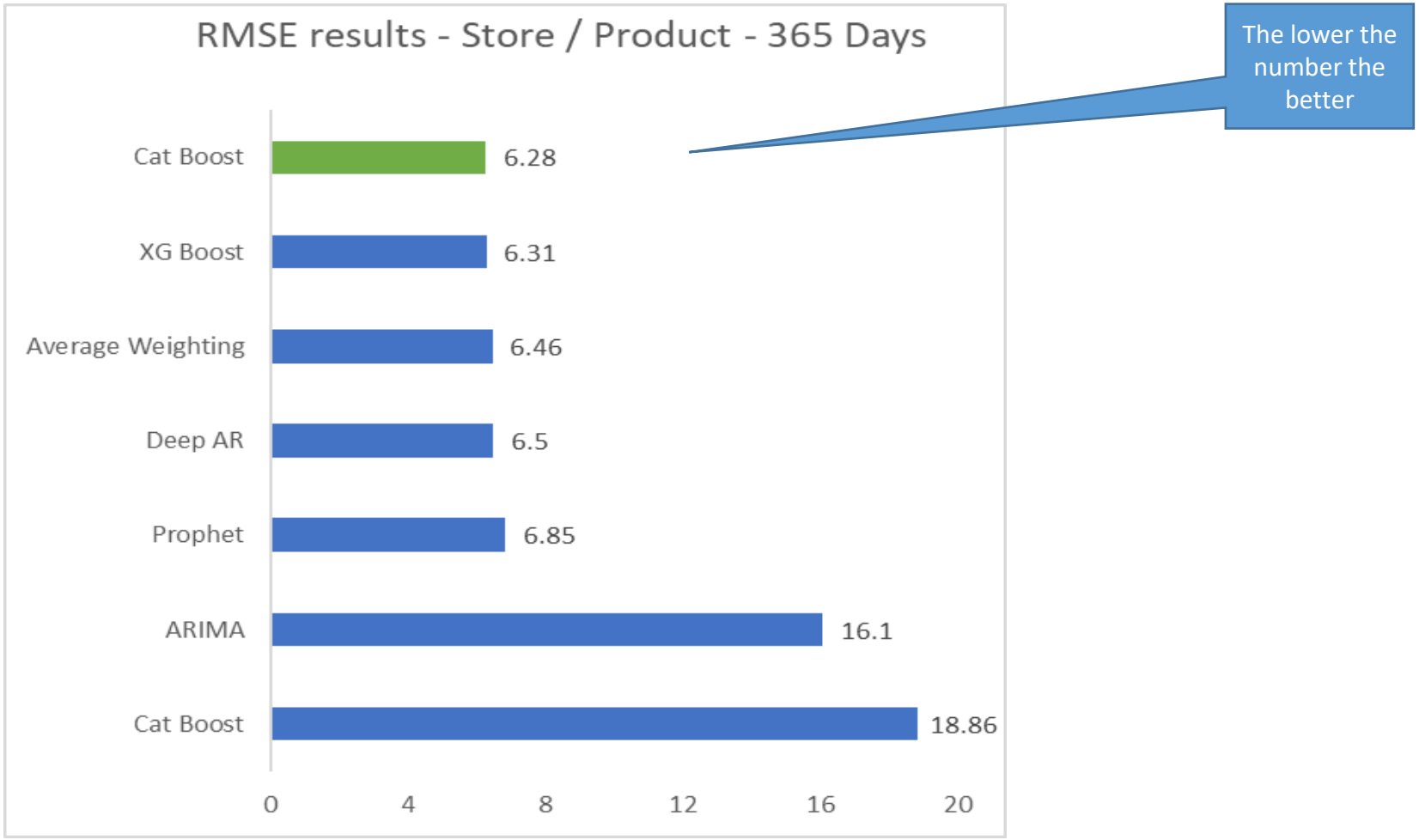
Grand Total

10,733,740 : actual sales
10,396,200 : forecast
3.1% : too low

Cat Boost was the best tool for by store by product over 365 days (500 forecasts)

Best

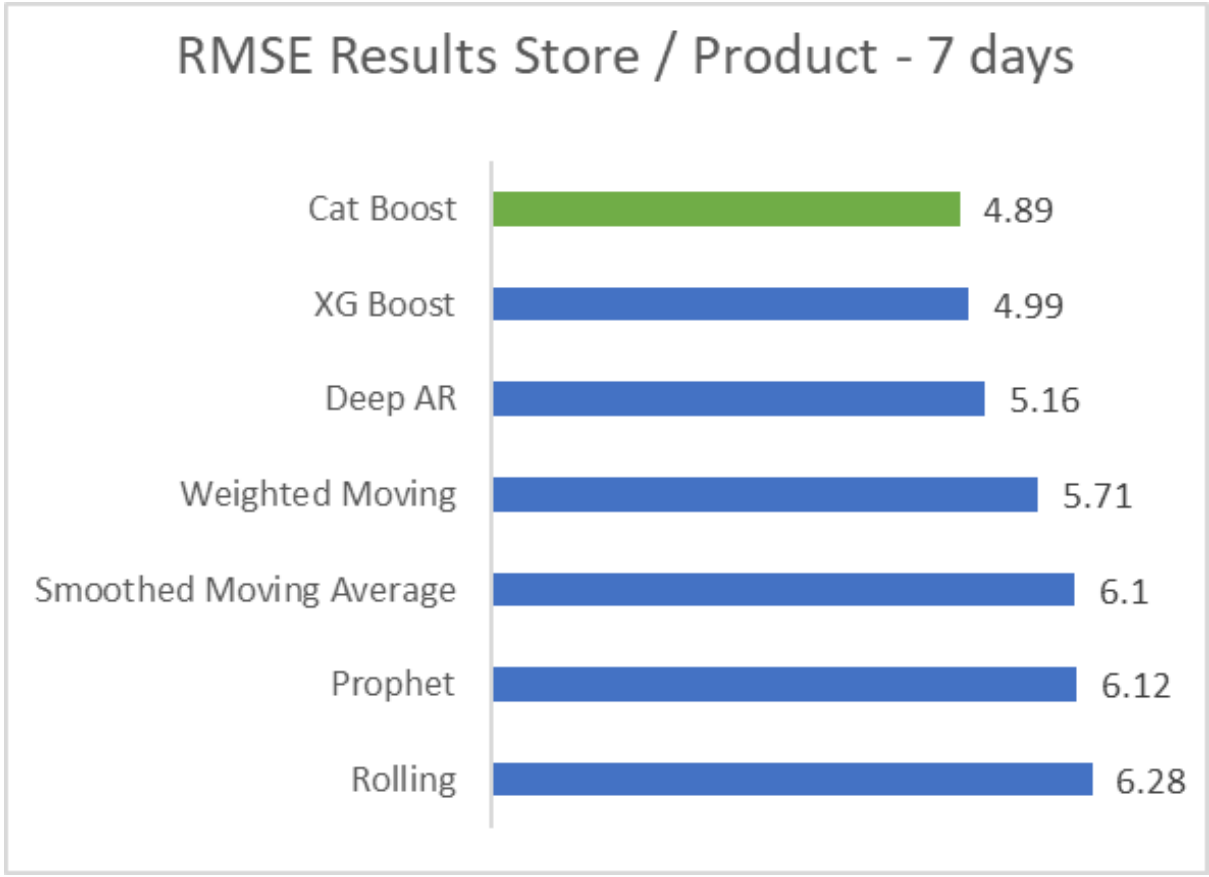
Other



Cat Boost was also the best for the 7 day time series solution vs rolling forecasts variations.

Best

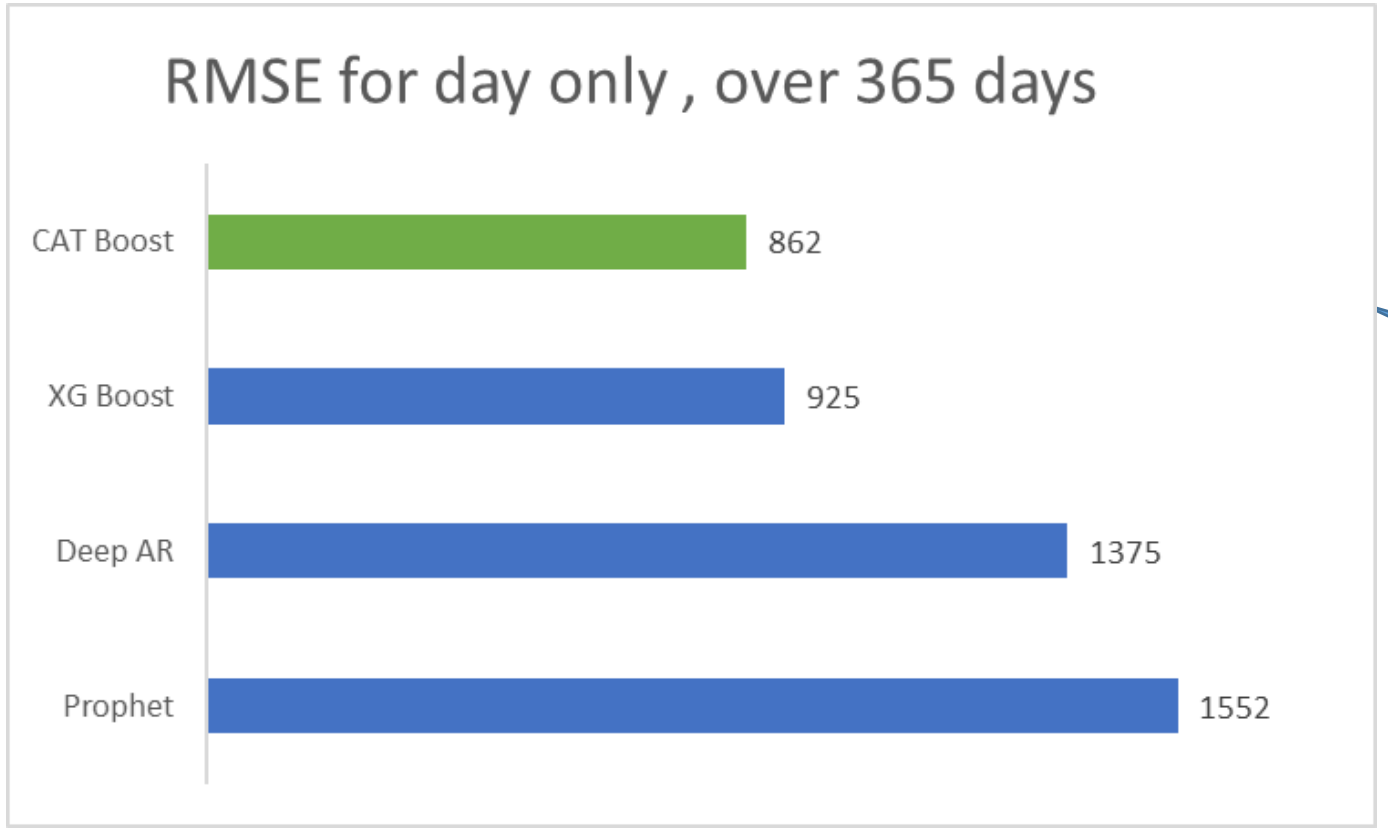
Other



Cat Boost was the best tool for daily forecasts

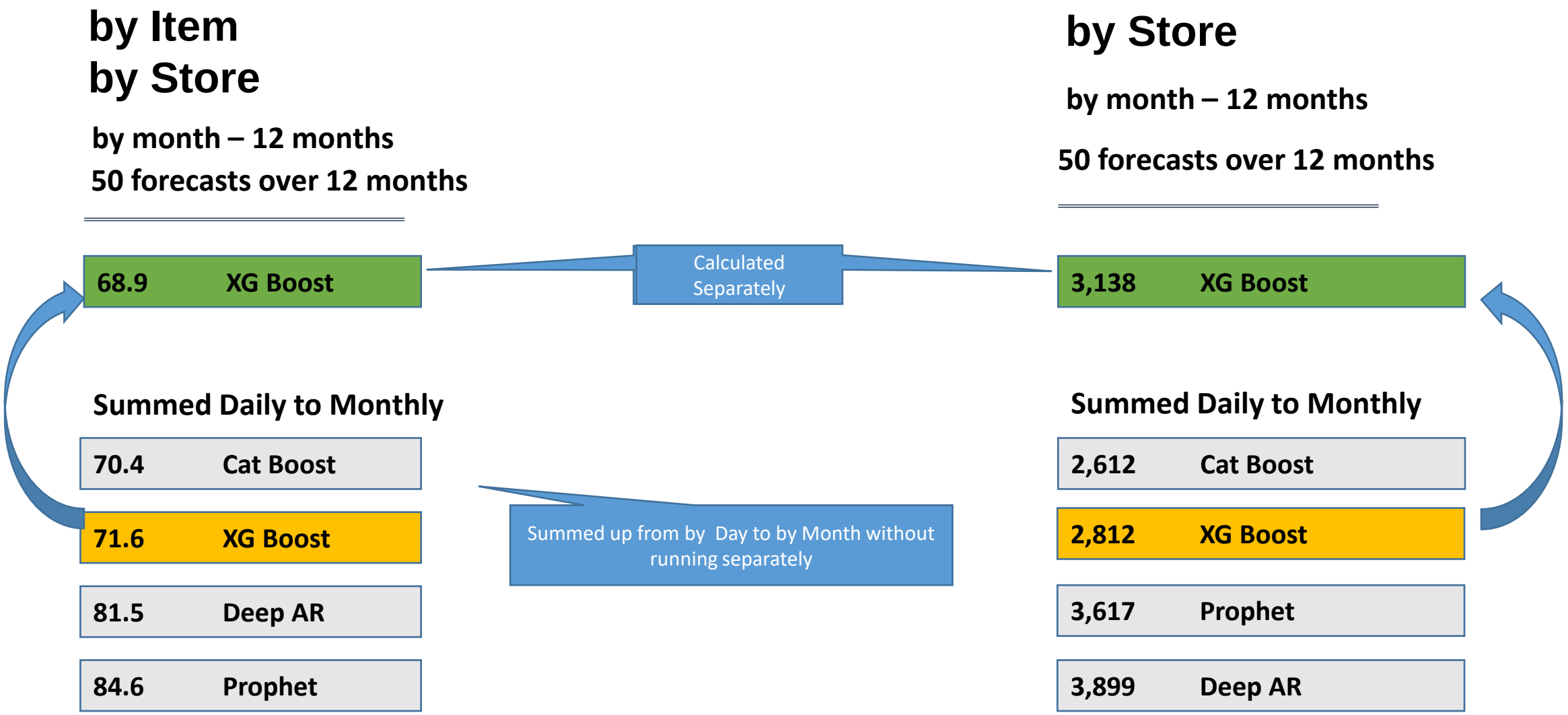
Best

Other



There were also other forecasts that had bad results that were not shown

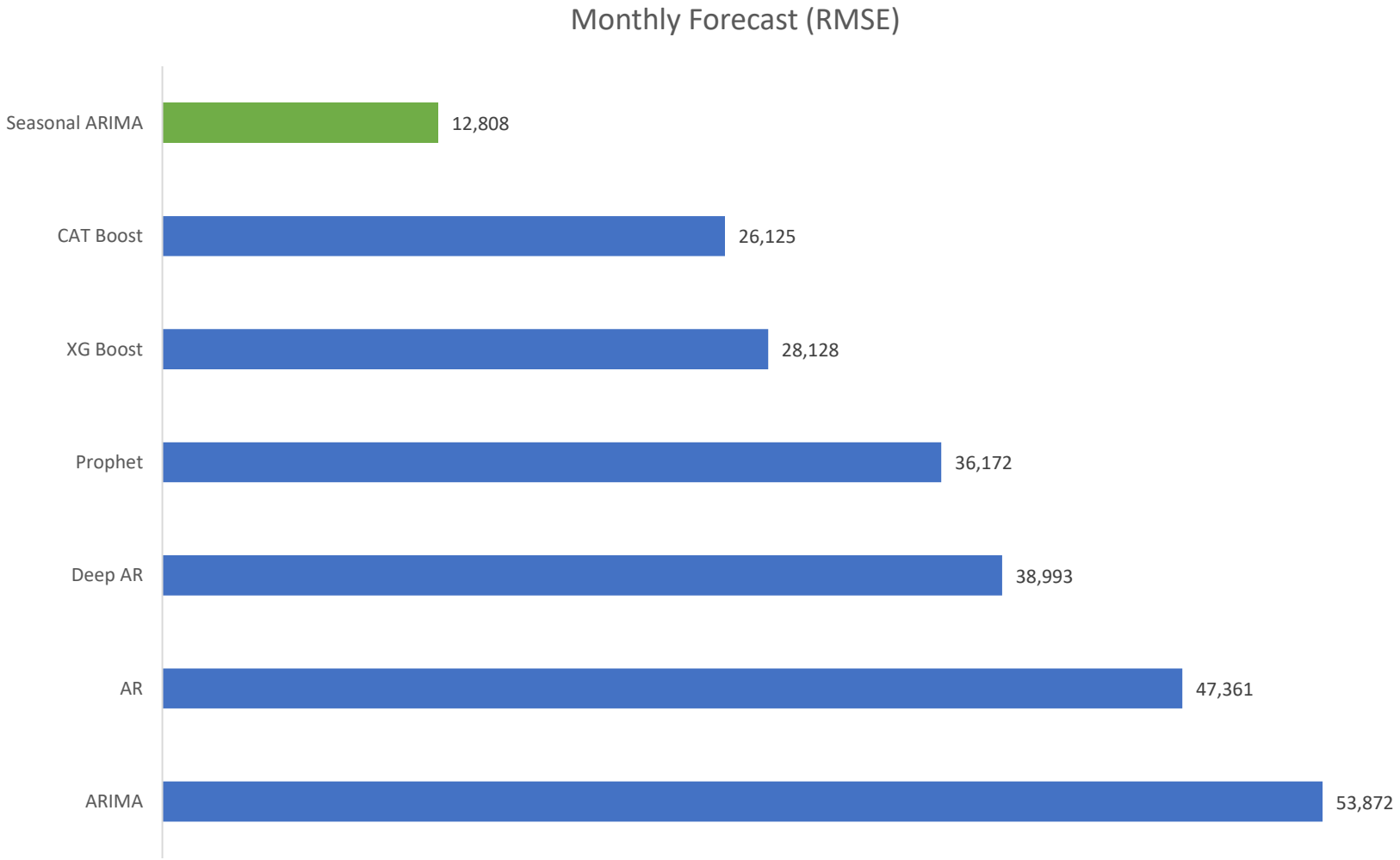
Can use the 1 forecast at store item and expand to other needs



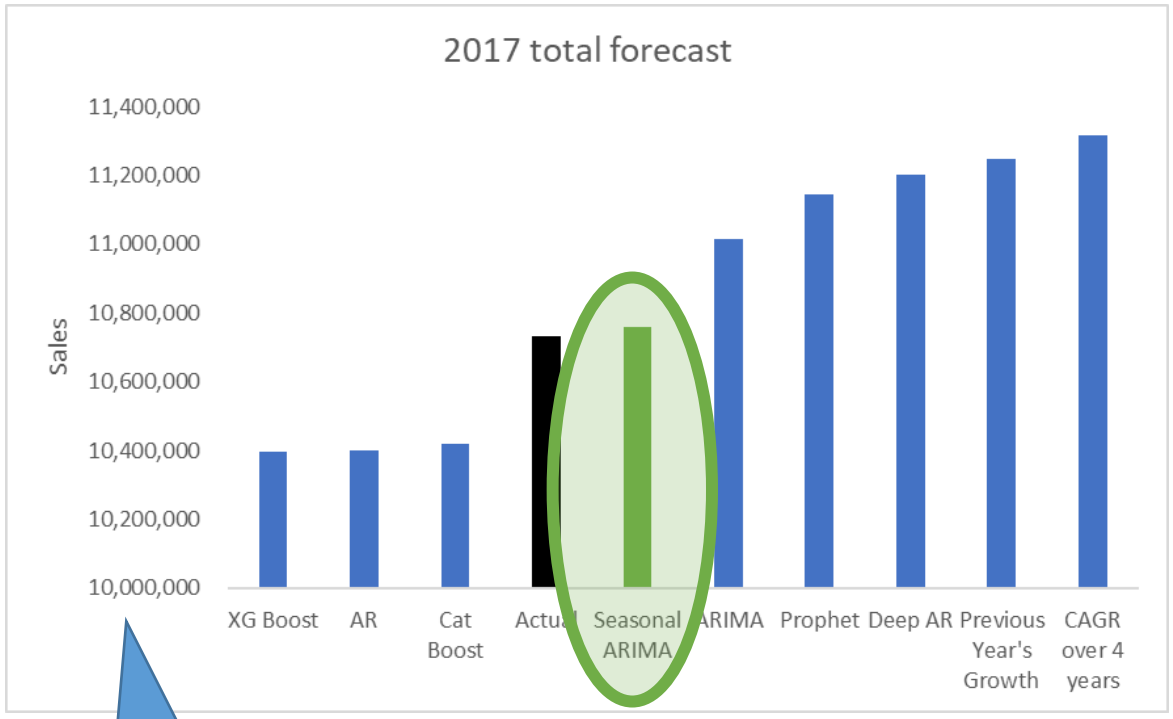
Seasonal ARIMA was the best for 1 forecast over 12 months

Best

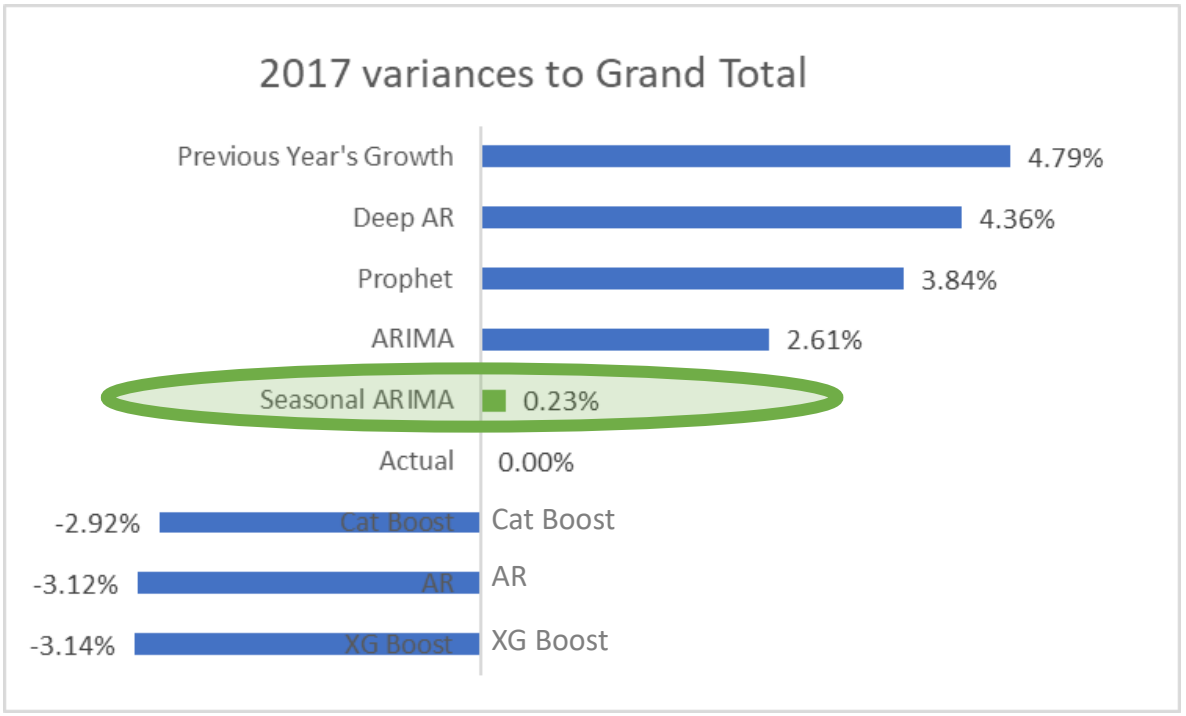
OK



Seasonal ARIMA was the best forecast for the Grand Total



Started scale at 10M



Many time series algorithms were completed to meet multiple business needs

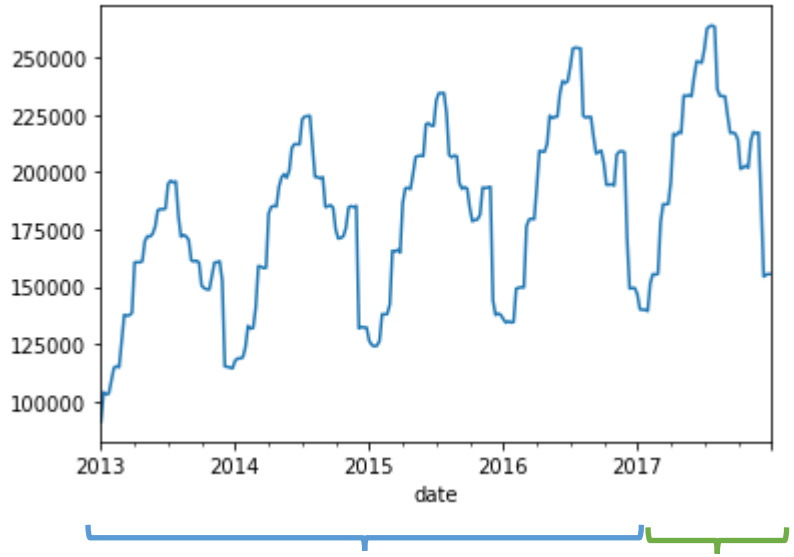
	By	Period	How Many forecasts	Units per forecast	Total forecast numbers
By Item by Store	Day	365 Days	7	500 (10 stores X 50 products)	1,227,500
By Item by Store	Day	7 Days	6	500	21000
By Day	Day	365 days	10	1	3,650
By Store	Month	12 months	4	10	480
By Item by Store	Month	12 months	4	500	2,000
By Month	Month	12 months	4	12	48
By Year	Year	1 year	10	1	10
		TOTAL	21		1,304,688

Further Options

Can scale up and use across other situations

10 stores → 1,000 stores → more stores

50 products → 5,000 products → more products



Can forecast by Week

Train on less time

Forecast on more time

COVID impact is training will often need to be shortened / adjusted.

i.e. April 2021 seasonality for will be different to April 2020 seasonality

Trained on 4 years of data

Forecasted and measured on 1 year of data

Thanks

Alex Dance

Background

- Maths / statistics degree
- Background in big data, strategy, analytics
- Worked at Optus, Salmat, Reuters, Pathfinder Solutions

Copy of This Presentation and code

https://github.com/alexdance2468/

Plus other data science projects completed

Contact Details

www.linkedin.com/in/alex-dance/

alexdance2468@gmail.com

0414 293 628