TEAM PROJECT

ARIMA ALGORITHM IN RETAIL

OUTLINE

OBJECTIVE

FORECASTING METHODS

TIME SERIES MODELS

CODE REVIEW

FINDINGS

CONCLUSIONS

OBJECTIVES

Concern about next year profit performance because company stock price is falling and the shareholders are not happy

Forecast next year performance to make sentimental better (if the forecast is good)

Assemble a competent team to do the task

Report to supervisor through Notion and presentation about the methods and findings

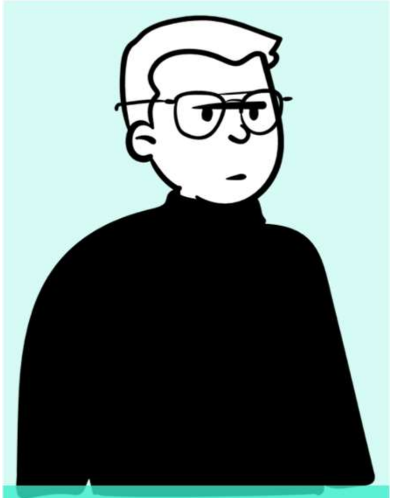
MEMBERS

Chris

Teamlead

Dzung

Data Scientist

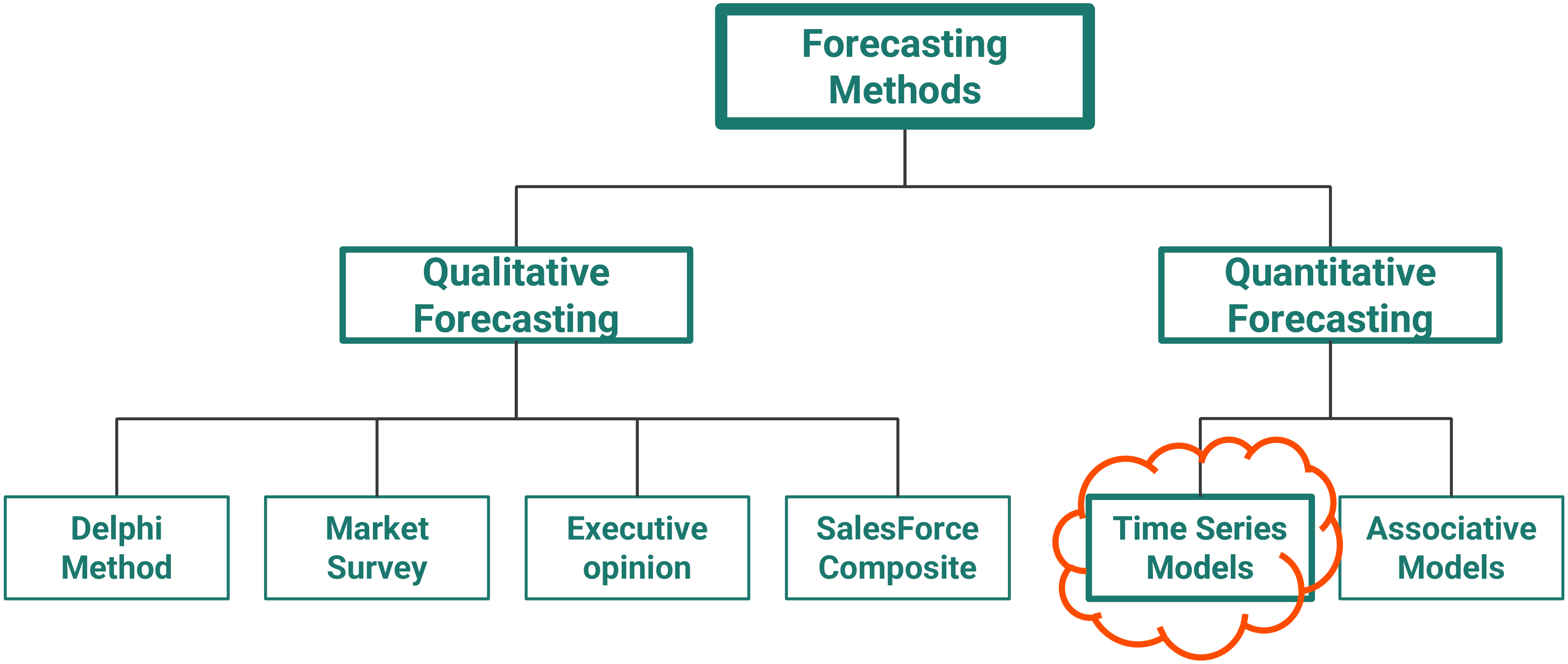


Jinbin

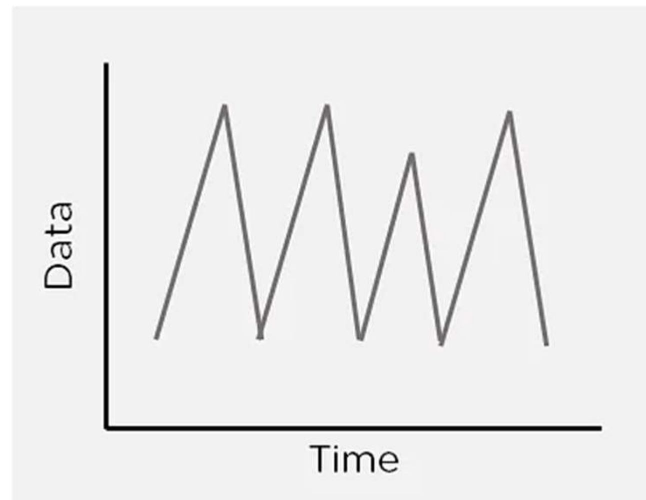
Data Scientist

” TIME SERIES FORECASTING “

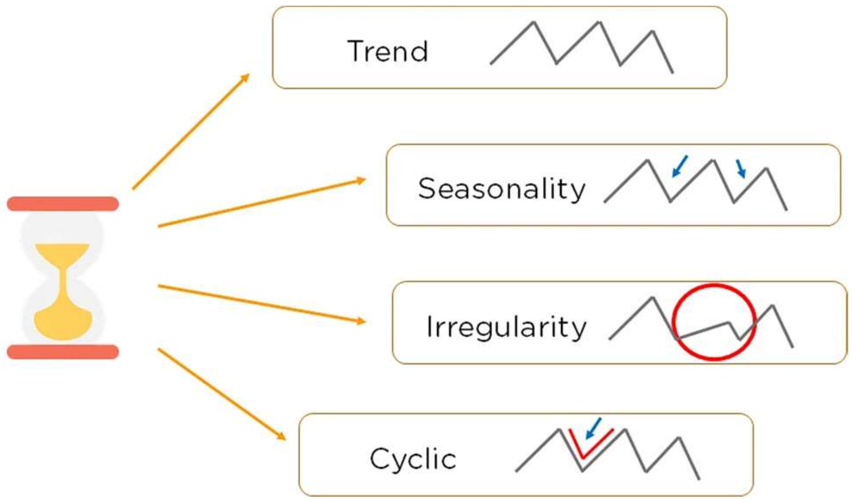
Forecasting Methods

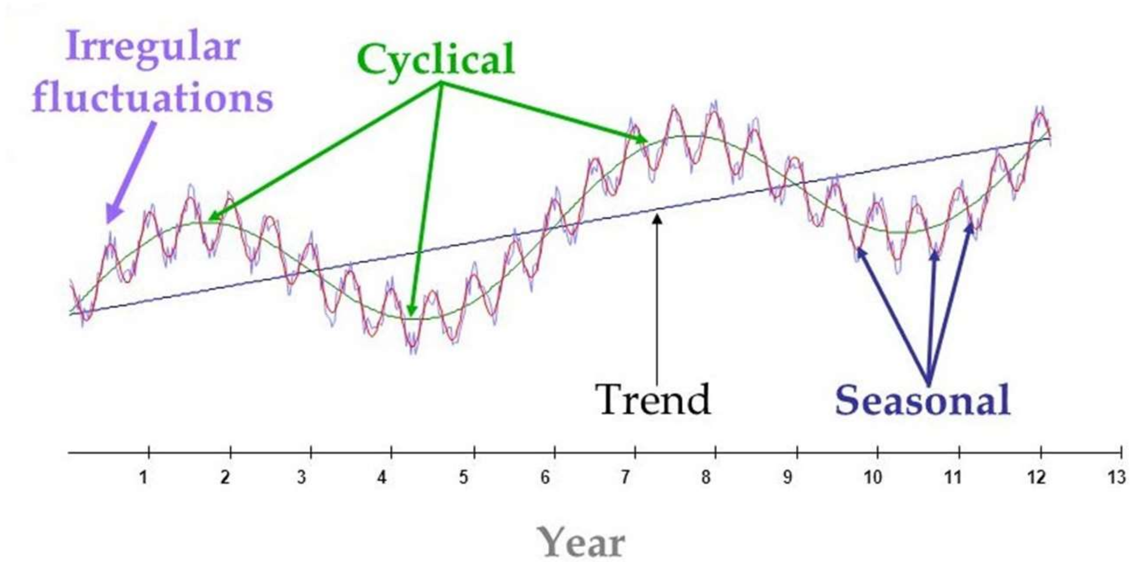


Time Series Models - Concept

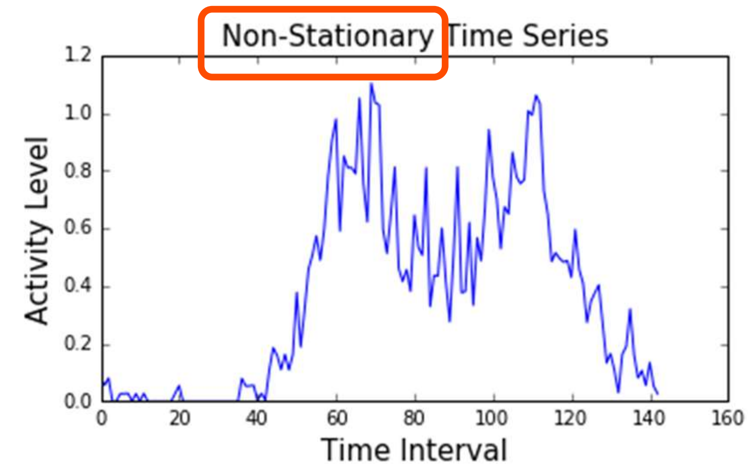


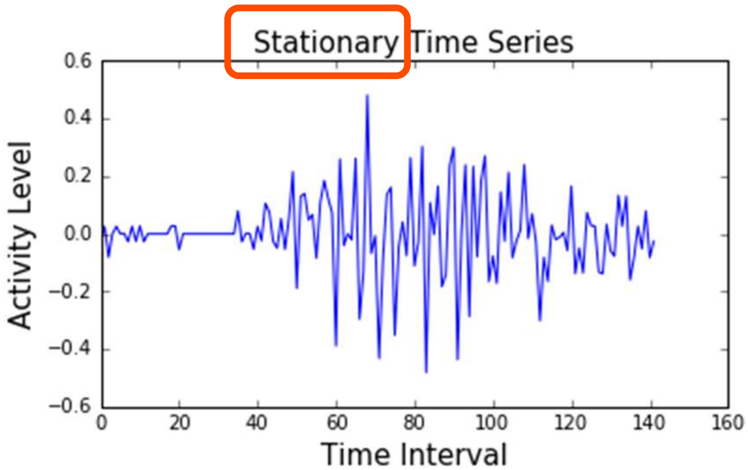
Time Series Models - Concept





Time Series Models - Concept



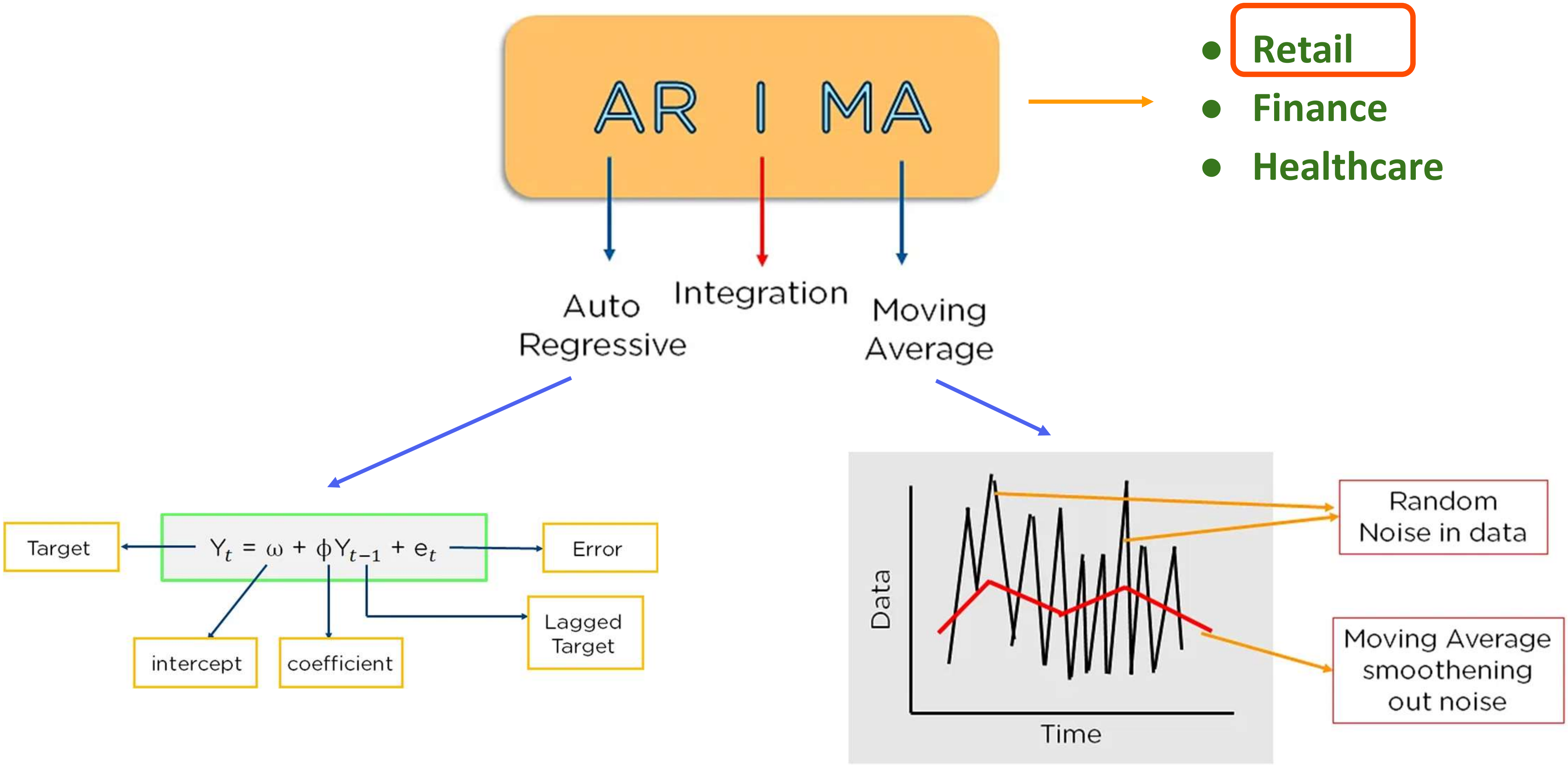


- Two types of time series data
- why Stationary?
- Check Stationary or not -> **ADF** (Augmented Dickey-Fuller) which is the most common test
- Non-Stationary -> apply some transformations

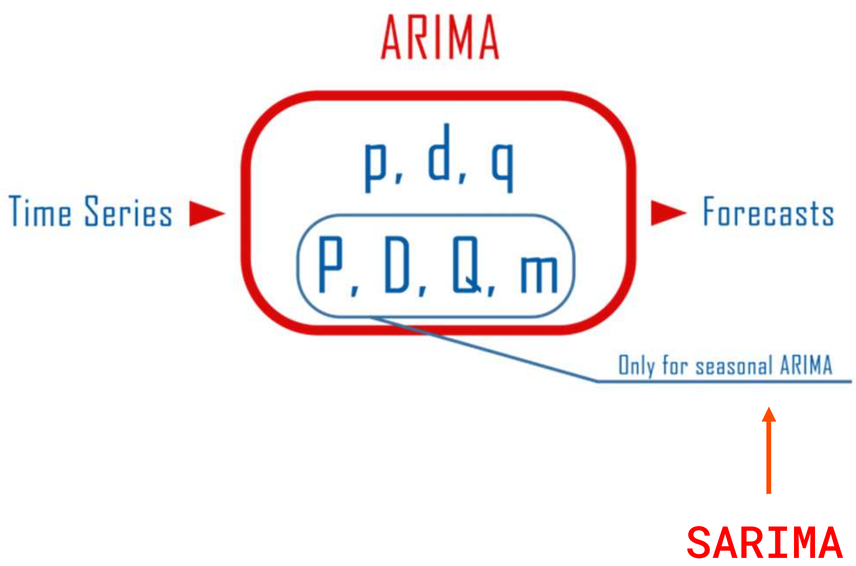
Time Series Models - Pros vs Cons

Model	Pros	Cons
Linear Regression	Ability to handle different time series components and features. High interpretability.	Sensitive to outliers. Strong assumptions.
Exponential Smoothing	Ability to handle variable level, trend, and seasonality components. Automated optimization.	Sensitive to outliers. Narrow confidence intervals.
ARIMA (Autoregressive Integrated Moving Average)	High interpretability. Realistic confidence intervals. Unbiased forecasts.	Requires more data. Strong restrictions and assumptions. Hard to automate.
Dynamic Linear Model	High interpretability. More transparent than other models. Deals well with uncertainty. Control the variance of the components.	Higher holdout error. Higher training and evaluation time.
Neural Network Model	Less restrictions and assumptions. Ability to handle complex nonlinear patterns. High predictive power. Can be easily automated.	Low interpretability. Difficult to derive confidence intervals for the forecasts. Requires more data.

Time Series Models - ARIMA



Time Series Models - ARIMA & SARIMA



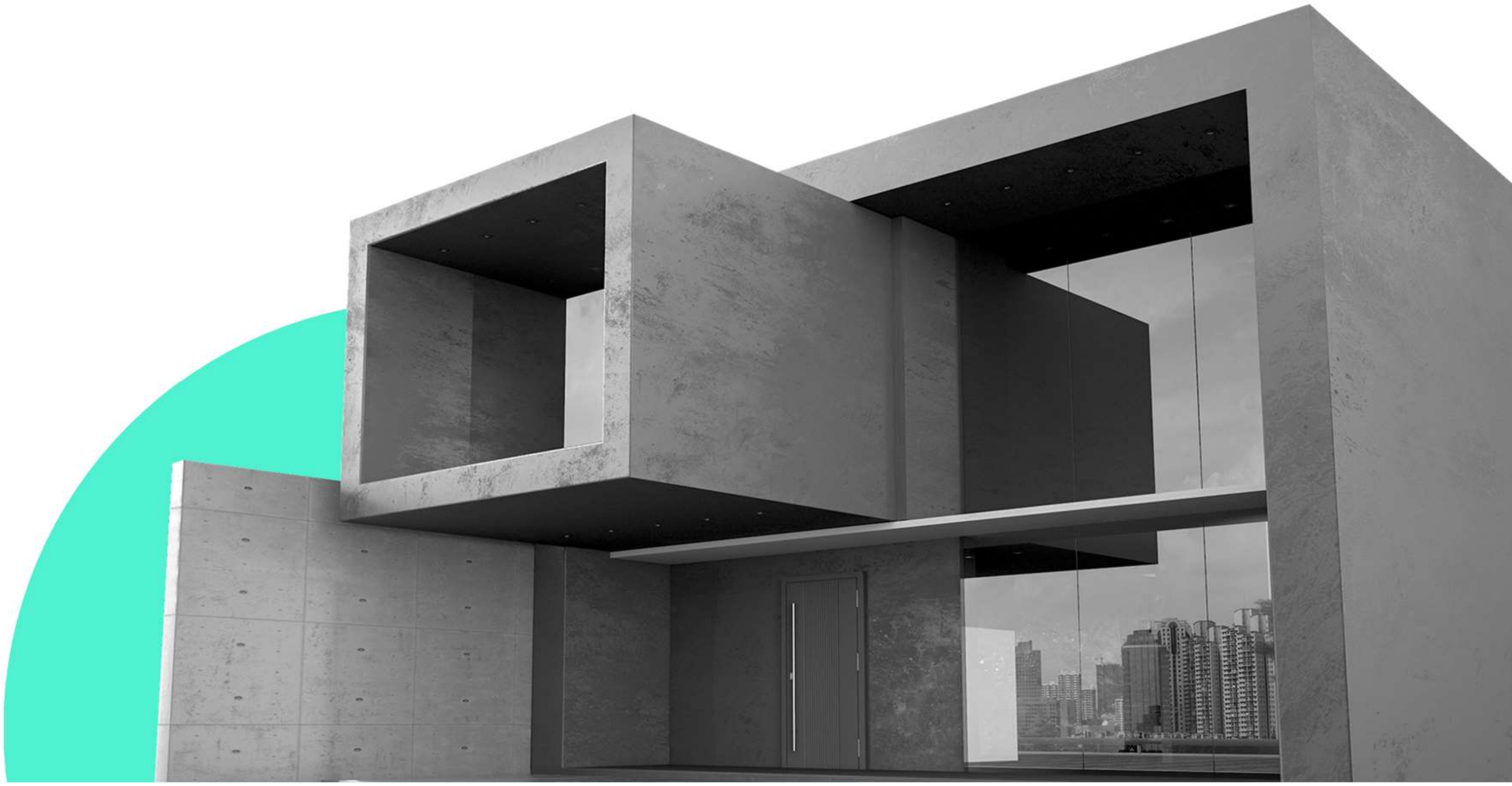
For **non-seasonal** data:

- **p**: Auto gressive (the no. of lag observations)
- **d**: Integration (the number of times that the raw differenced observations)
- **q**: Moving Average (the size of the moving average window)

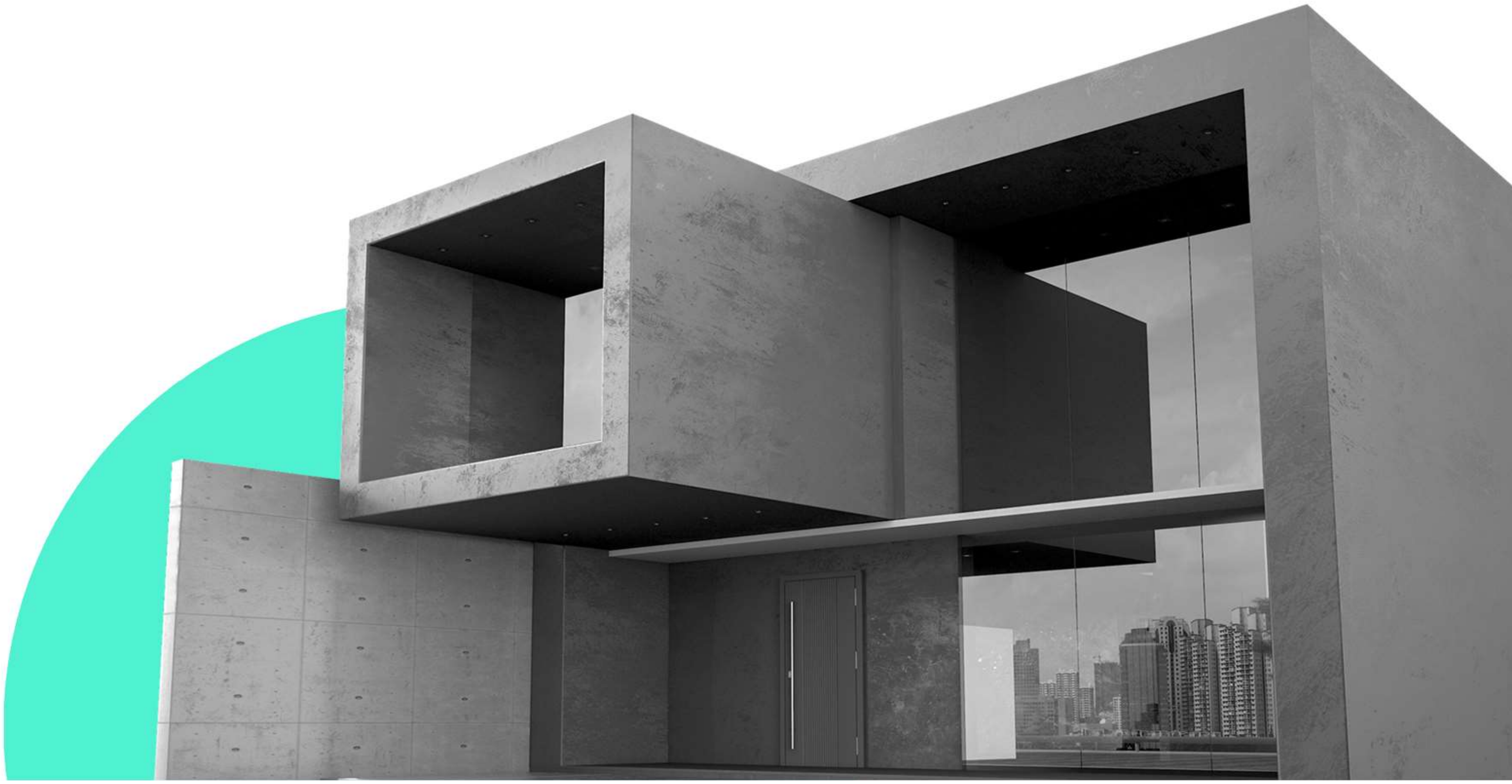
For **seasonal** data:

- **P**: The number of seasonal lag observations the model
- **D**: The number of times that the differenced seasonal observations
- **Q**: The size of the seasonal moving average window
- **m**: The number of observations of 1 season

CODE REVIEW



FINDINGS



TIME SERIES MODELS - Model Evaluation

sklearn.metrics.mean_absolute_percentage_error

`sklearn.metrics.mean_absolute_percentage_error`(*y_true, y_pred, *, sample_weight=None, multioutput='uniform_average'*)

[source]

Mean absolute percentage error (MAPE) regression loss.

$MAPE$	Forecasting power
<10%	Highly accurate forecasting
10%~20%	Good forecasting
20%~50%	Reasonable forecasting
>50%	Weak and inaccurate forecasting

Source: Lewis (1982)

```
from  sklearn.metrics import mean_absolute_percentage_error

mape= mean_absolute_percentage_error(test, pred)

print('MAPE: %f' %mape)
```

```
MAPE: 0.363205
```

Forecast result

Month/Year	2018(Predicted)
1	-34.242371
2	59.325691
3	25.207121
4	43.463854
5	45.450554
6	19.324991
7	31.822628
8	15.33349
9	50.541756
10	15.778343
11	50.901949
12	56.173248

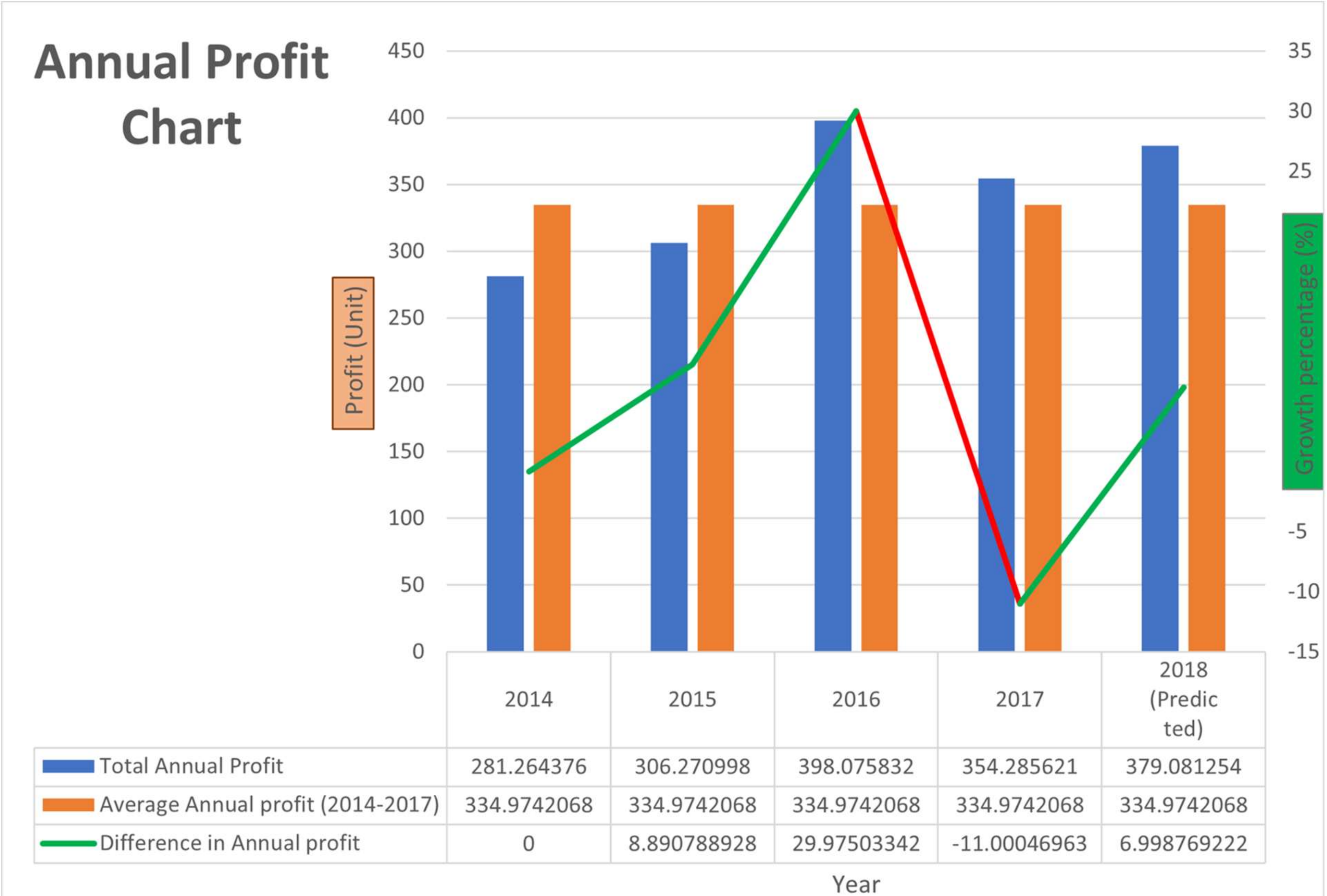
Worst performance : January (-34.24 Unit))

Best performance: February (+59.32 Unit)

Average (mean) performance: +31.5 Unit

Comparing forecasted result

Month/Year	2014	2015	2016	2017	2018(Predicted)
1	31.01507	-56.5691	31.73959	46.06735	-34.242371
2	18.74584	43.96642	60.29614	15.08292	59.325691
3	3.176624	70.52245	22.15931	61.98274	25.207121
4	25.84322	26.17185	17.51656	4.597488	43.463854
5	22.44844	31.97171	38.49843	26.20902	45.450554
6	36.86314	24.1707	23.87125	33.56464	19.324991
7	-5.88449	23.49035	22.05412	30.76381	31.822628
8	34.75886	33.68433	11.7163	41.47227	15.33349
9	31.075	28.01762	25.69878	23.94674	50.541756
10	21.68715	16.97208	82.87318	31.12509	15.778343
11	29.22053	38.50243	10.84164	21.11134	50.901949
12	32.315	25.37015	50.81054	18.36222	56.173248



	2014	2015	2016	2017	2018 (Predicted)
Total Annual Profit	281.264376	306.270998	398.075832	354.285621	379.081254
Average Annual profit (2014-2017)	334.9742068	334.9742068	334.9742068	334.9742068	334.9742068
Difference in Annual profit	0	8.890788928	29.97503342	-11.00046963	6.998769222

Conclusions and Recommendations

Overall forecast for next year is good!

Share it with shareholder and public to make them feel comfortable

Keep investing in new opportunities without worry too much about survival

Bad January 2018 profit !

Even though the forecast for next year is good but for January is really bad, it could resemble the year 2015

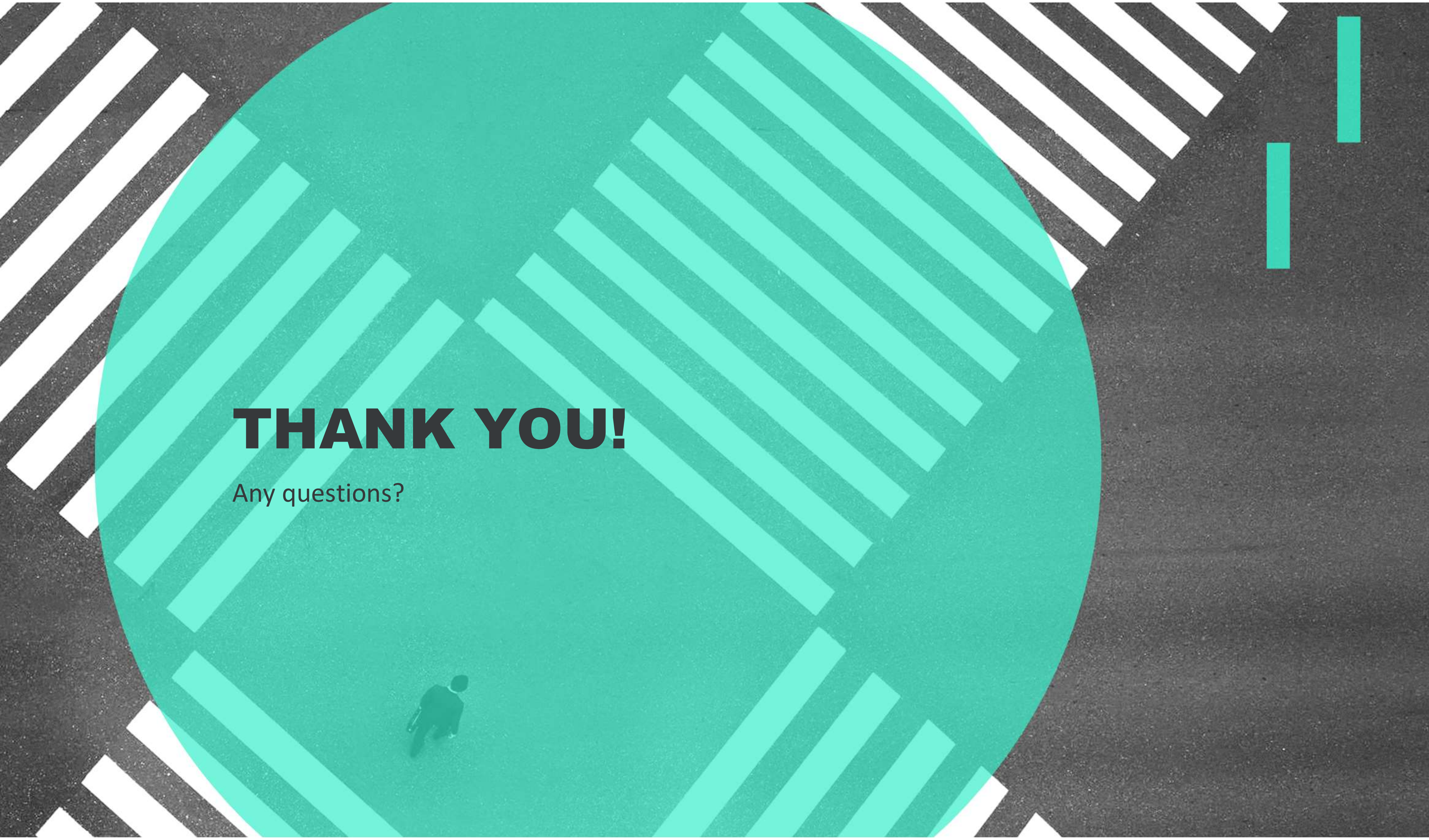
Investigate what did happen in that year so we can avoid it if it happens again

Small Data team

Invest more in better data engineering to have quality dataset

Expand the data team to better the workflow

Improve computational power to satisfy bigger data



THANK YOU!

Any questions?