

Detecting Project Overruns with Time Series Models

Introduction



Zühlke is a global innovation service provider.

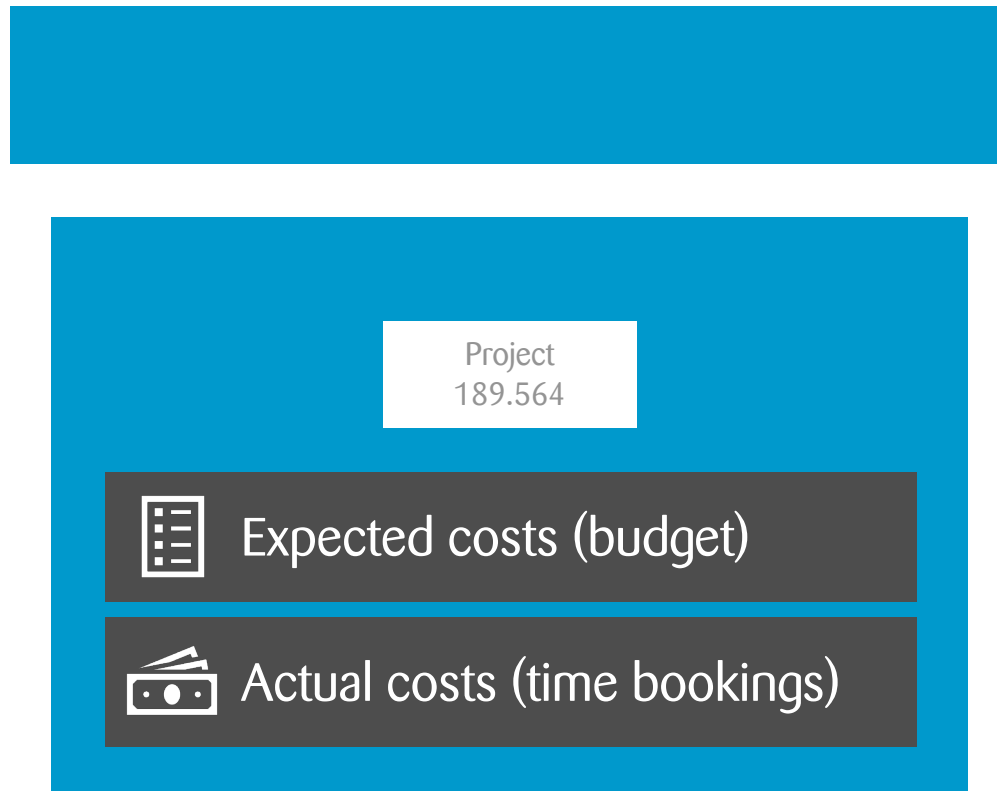
As such, we realize many projects.



Introduction



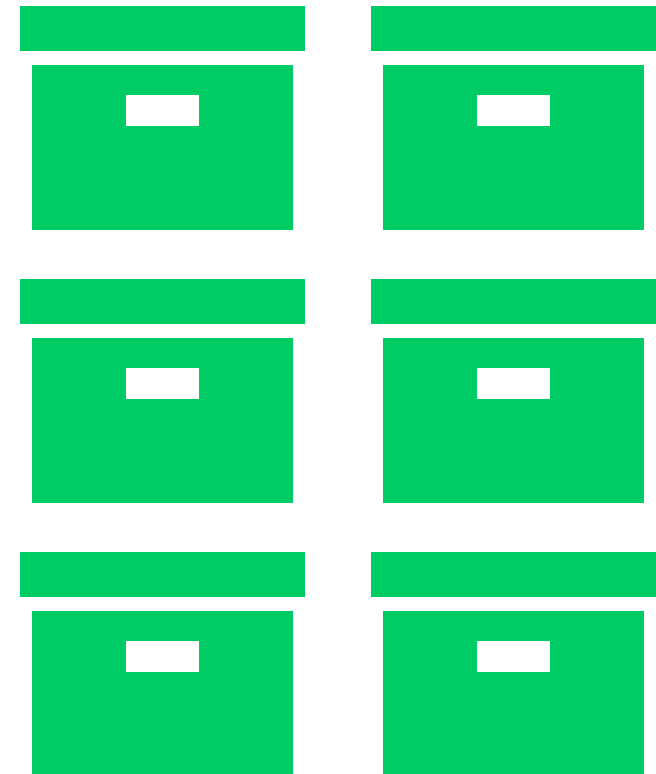
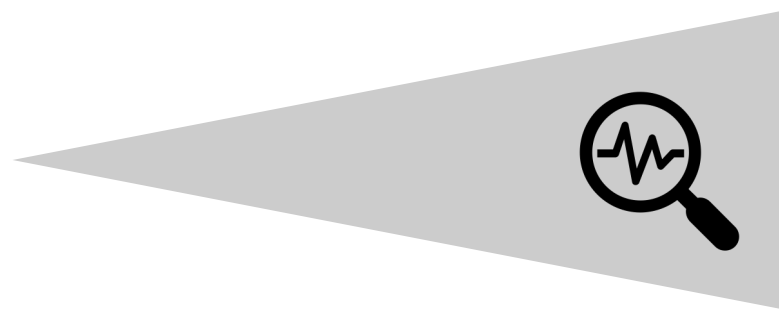
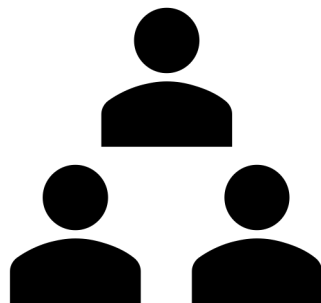
A project consists of expected and actual costs.



Introduction



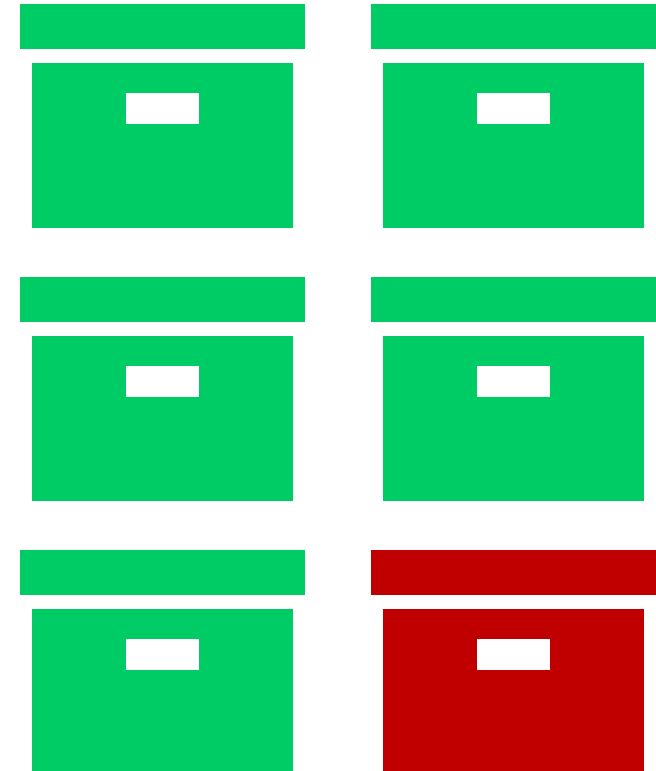
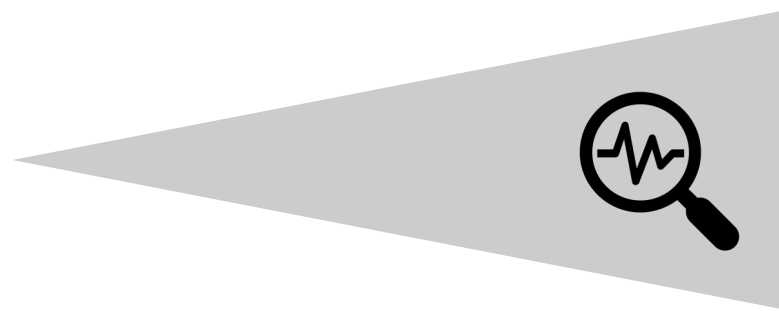
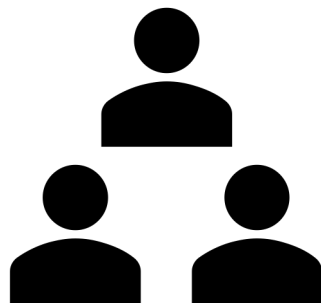
PM and a dedicated team control these.



Introduction



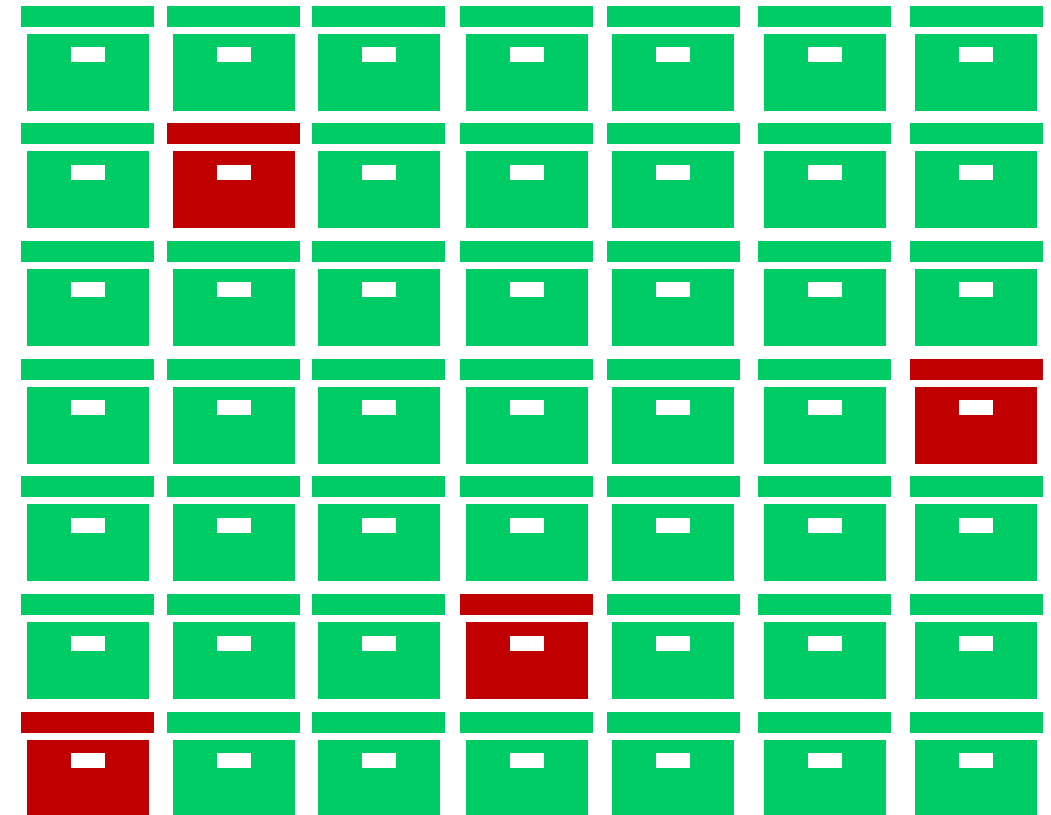
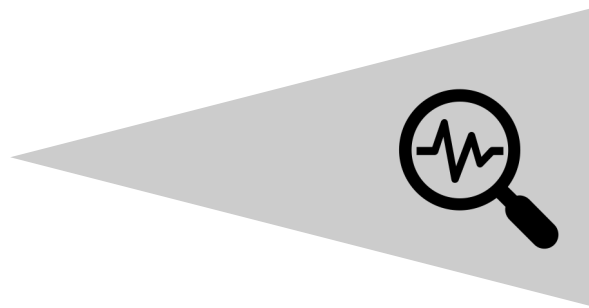
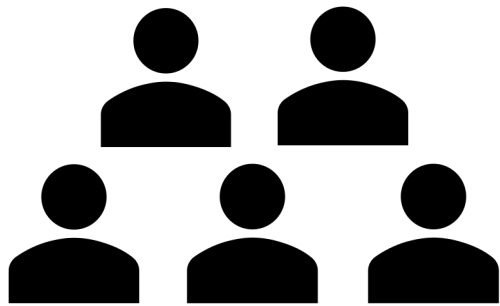
Project overruns should be detected as early as possible.



Introduction



With a growing number of projects however, this becomes increasingly challenging.



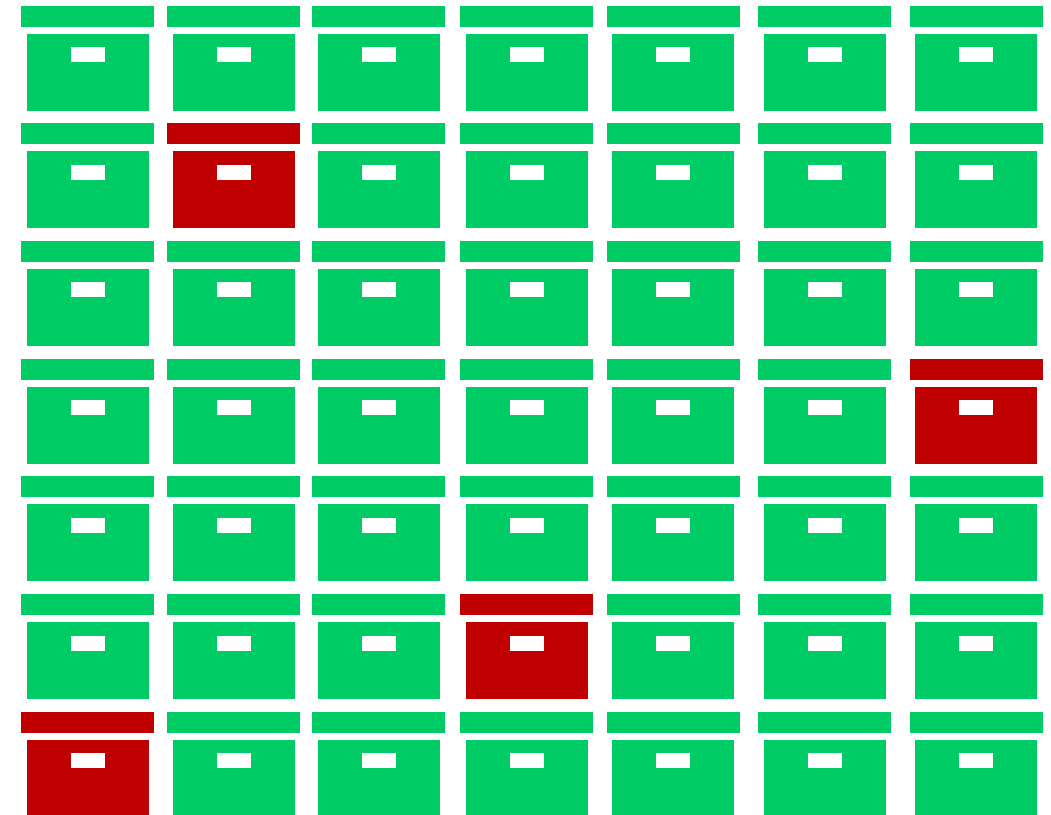
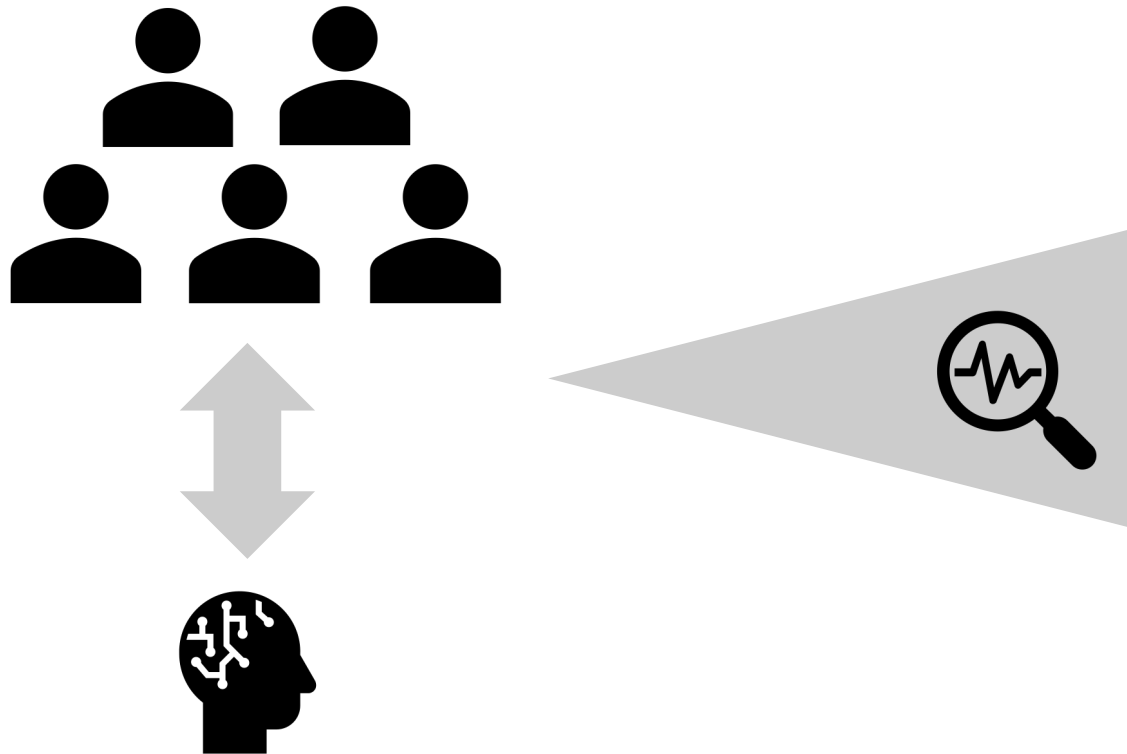
Zühlke 2021:

- 1'500 employees
- 500+ concurrently running projects
- Thousands of concurrently running project phases

Introduction



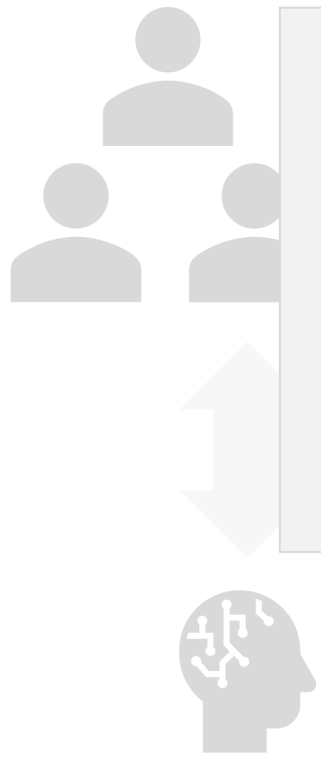
ML might be able to support us in this task.



Introduction



ML might be able to support us in this task.

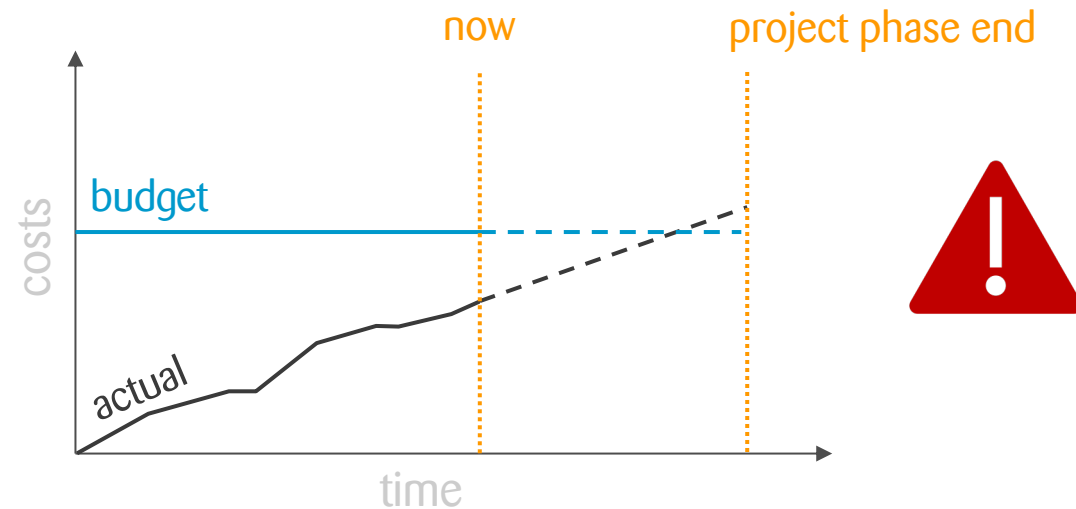


Z-POP
Zühlke Project Overrun Prediction



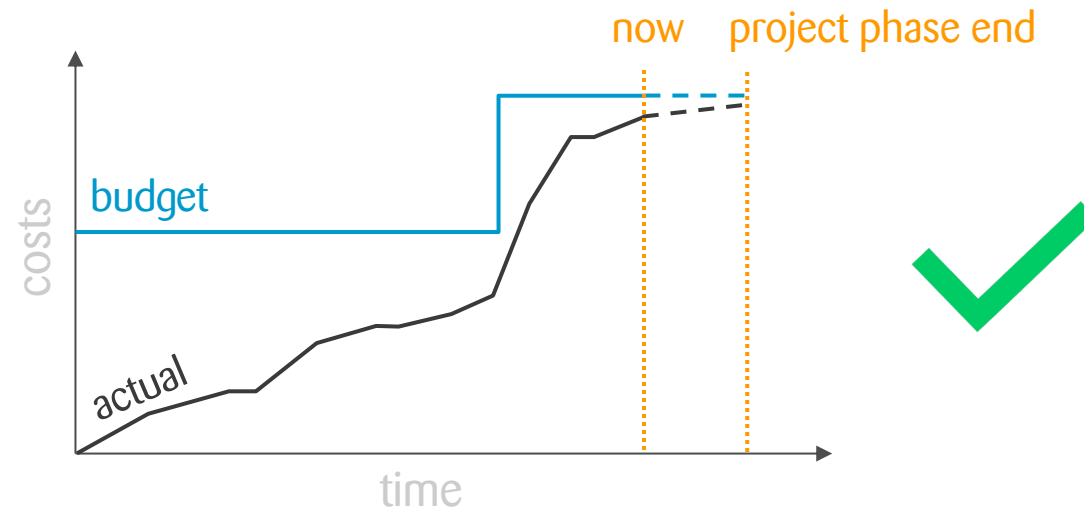
Problem Modeling

- Project members book time onto a project phase
- These costs result in time series data
- The time series can be forecast and compared to the budget



Problem Modeling

- Project members book time onto a project phase
- These costs result in time series data
- The time series can be forecast and compared to the budget



Dataset



- To evaluate our methods, we exported 1 year of data from our systems
- We then selected a subset of the projects based on duration, minimum budget, etc.



Approach: Baseline

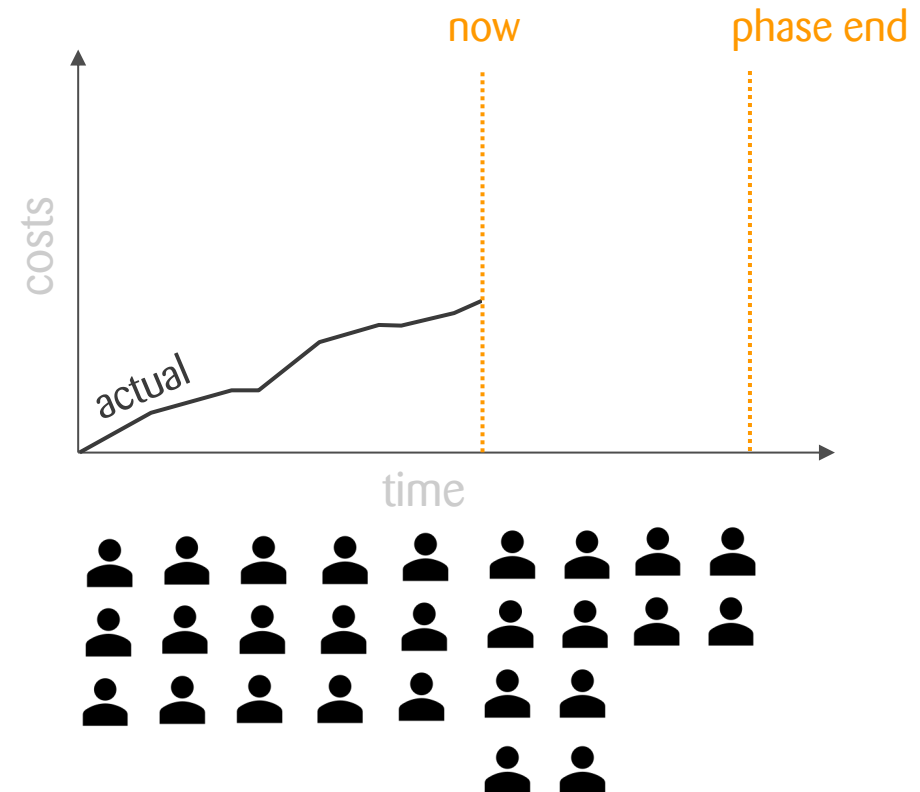


Project 189.564

Staffing Info

Expected costs (budget)

Actual costs (time bookings)



Approach: Baseline

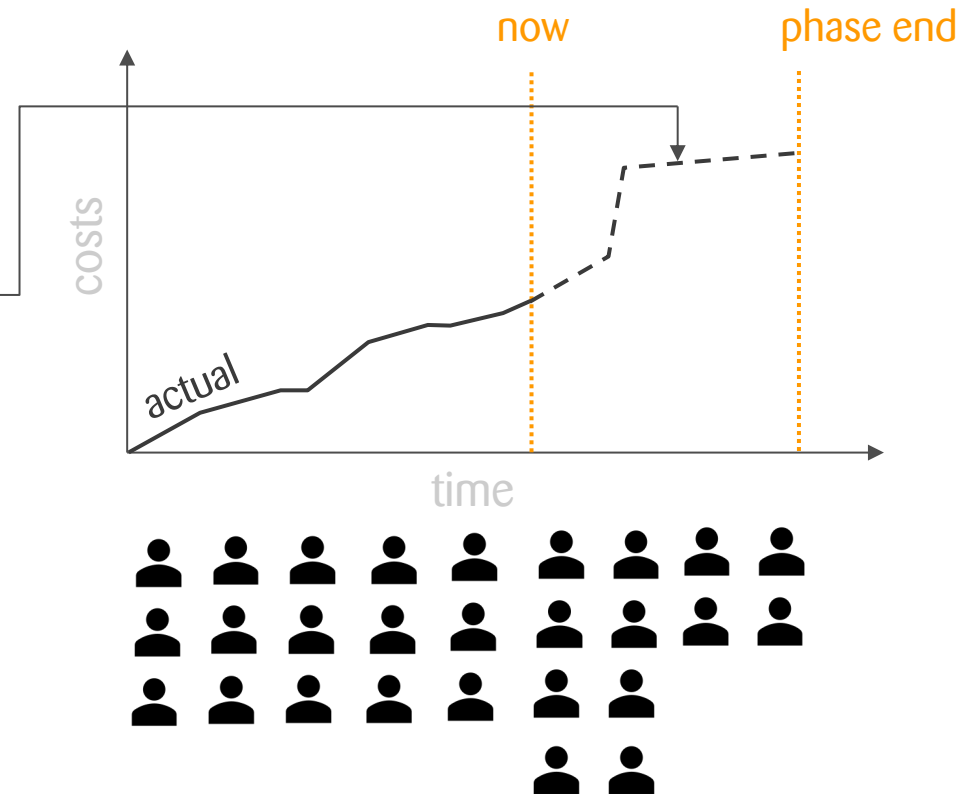


Project 189.564

Staffing Info

Expected costs (budget)

Actual costs (time bookings)



Evaluation: Regression



rel. MAE
after 75% of phase duration

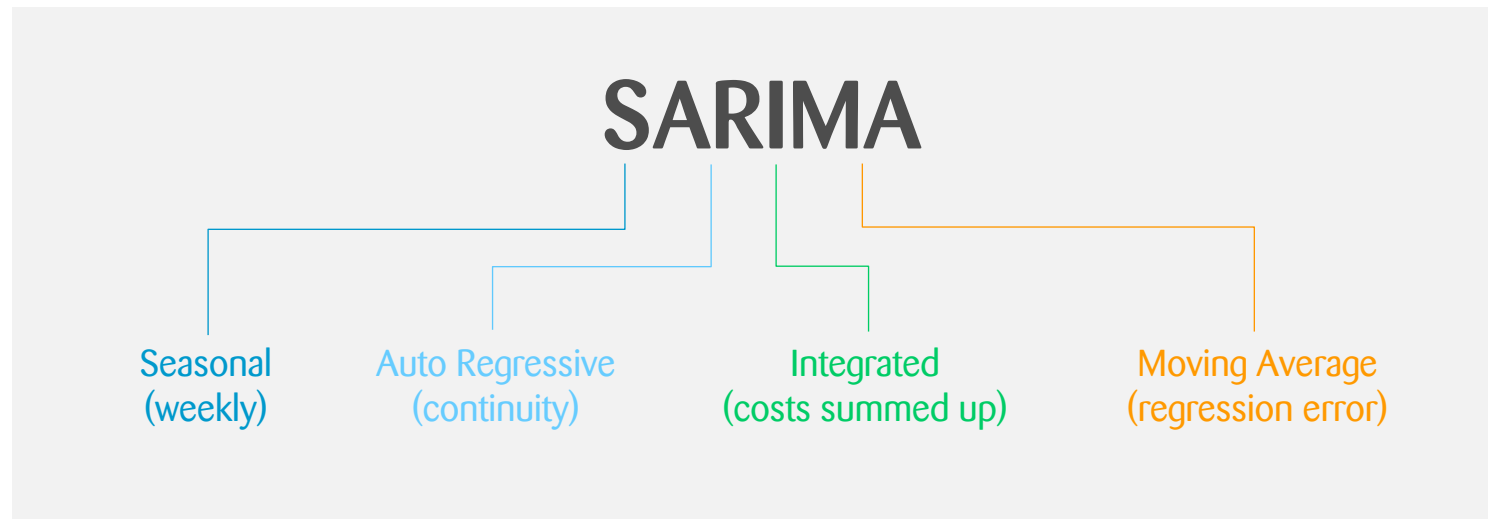
Baseline

26%

Approach: SARIMA



Weakness Baseline: Does not consider previous costs



- Popular choice in timeseries analysis and forecasting
- We found that order can be fixed for all cost series

Evaluation: Regression



rel. MAE
after 75% of phase duration

Baseline

26%

SARIMA

21%

Approach: SARIMAX



Weakness SARIMA: Does not consider planned costs

Baseline + SARIMA =
SARIMAX

eXogenous factors
(baseline)

- Add baseline as exogeneous factor into SARIMA to get best of both worlds

Evaluation: Regression



rel. MAE
after 75% of phase duration

Baseline	26%
SARIMA	21%
SARIMAX	16%

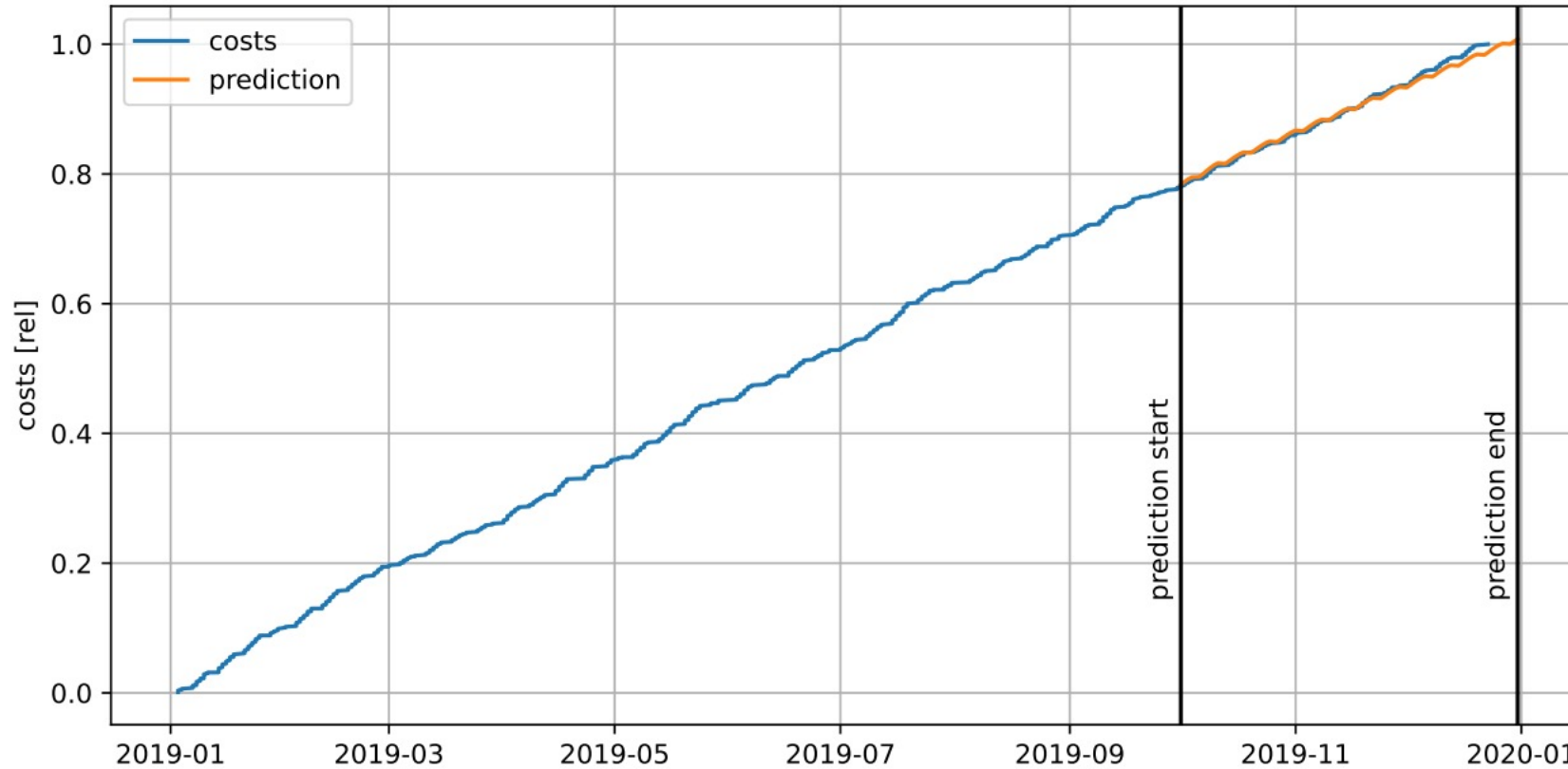
Evaluation: Classification



- Use forecast to detect overruns
- Define different thresholds (yellow, orange, red) depending on predicted overshoot

	green	yellow	orange	red	
no overrun	137	10	7	4	
overrun	6	11	8	6	
		81%	45%	19%	recall
		54%	56%	60%	precision

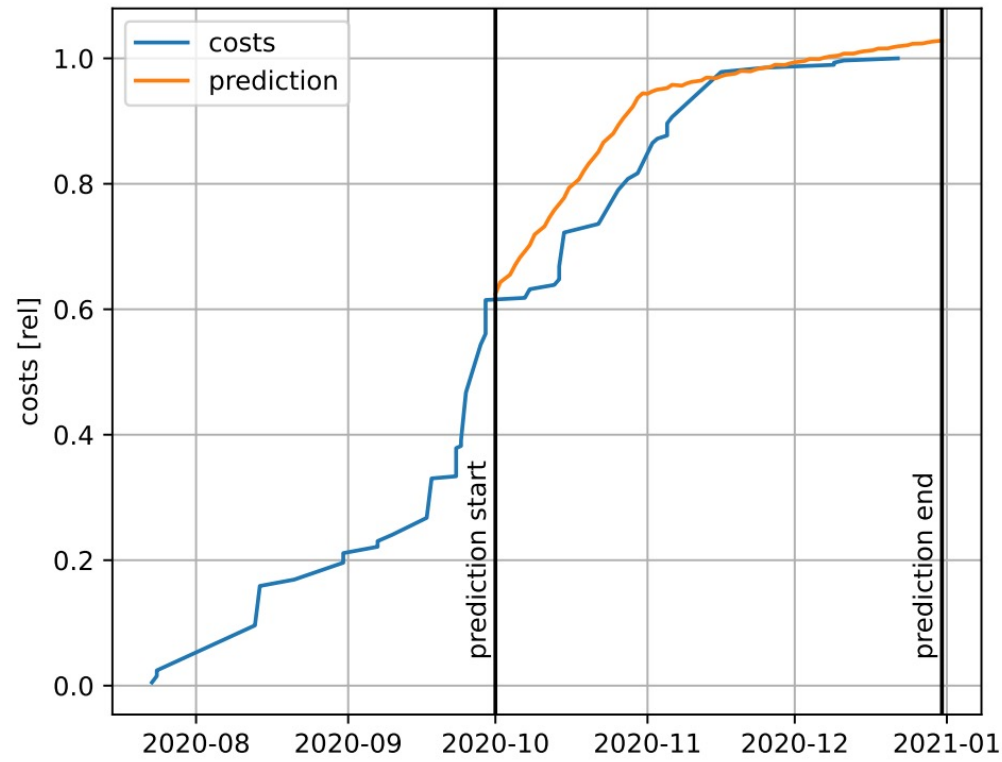
Example: Simple Case



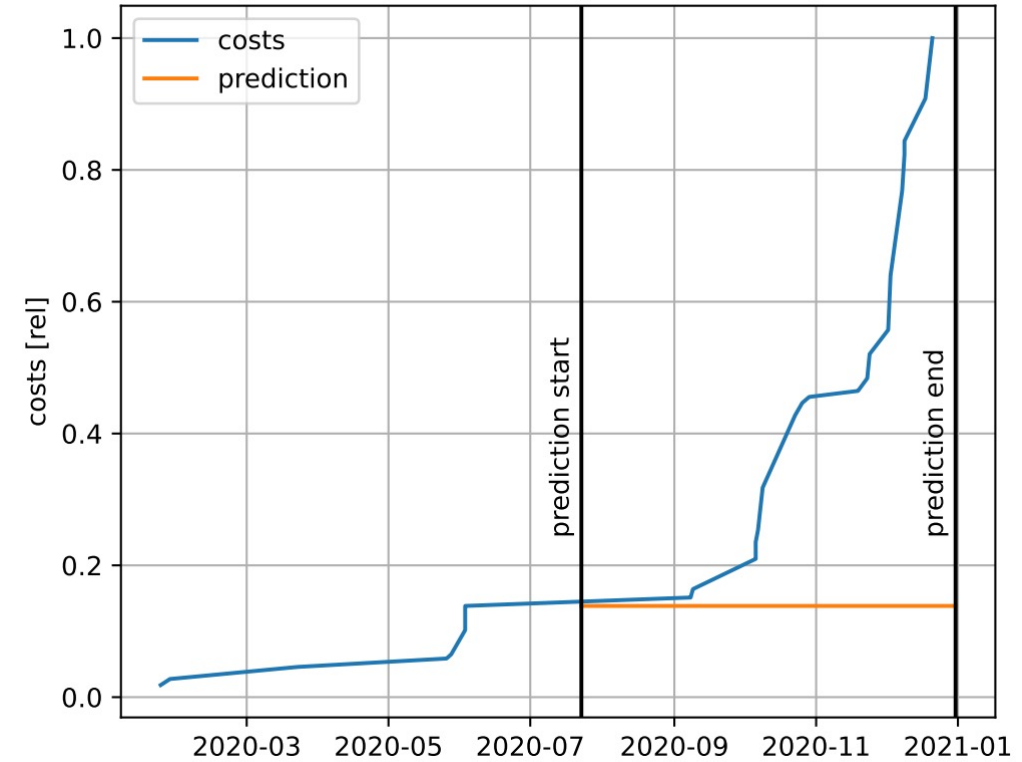
Examples: Complex Cases



with useful staffing info



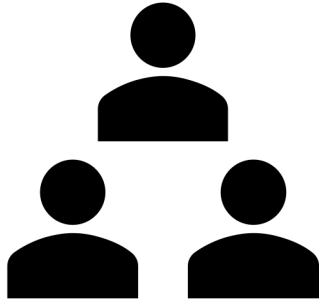
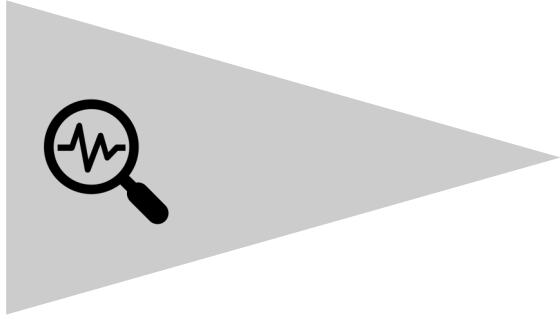
without staffing info



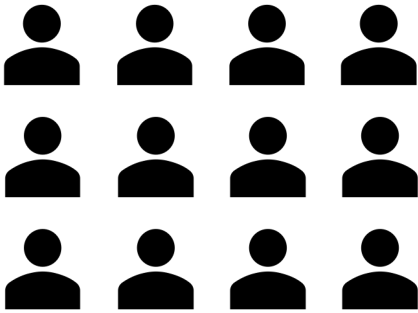
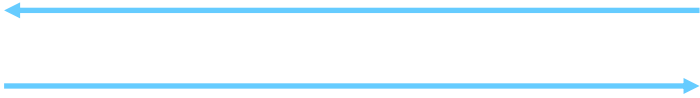
Integration



Project 1845-329	●
Project 4594-378	●
Project 1987-452	●
Project 5894-127	●
Project 6894-516	●
Project 8973-654	○
Project 2764-986	●
Project 1327-198	●



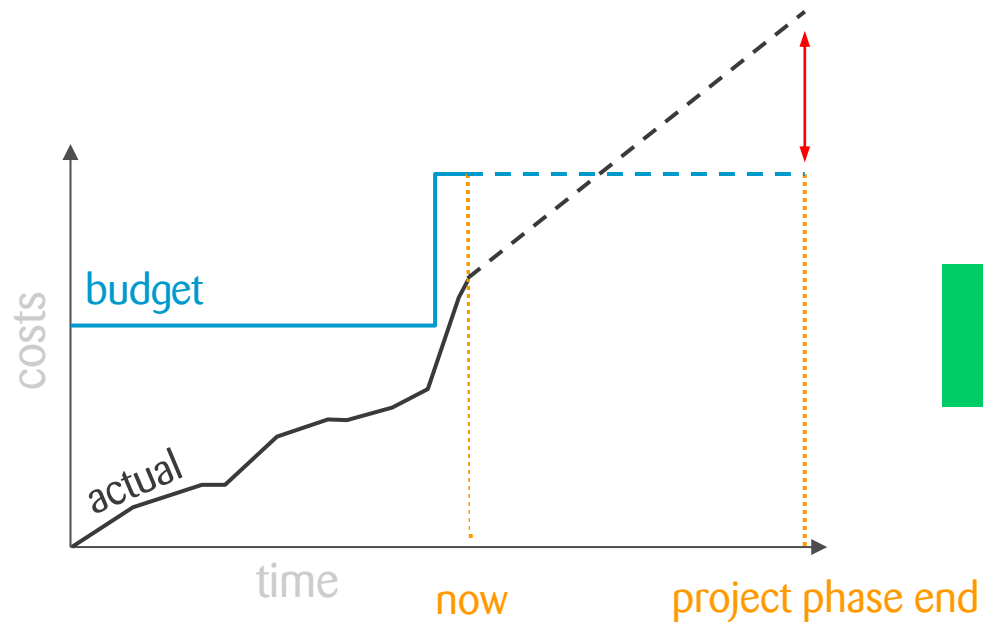
Z-POP



Findings



■ Data Quality



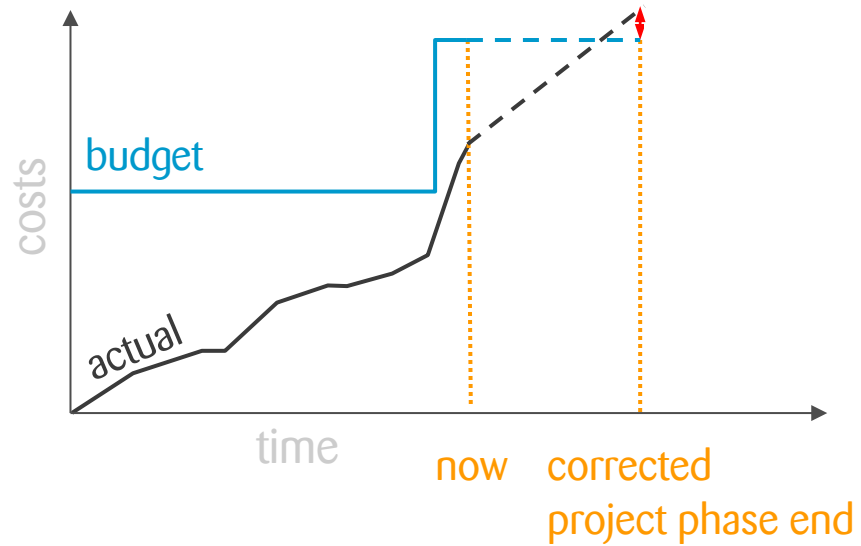
SARIMAX

16%

Findings



■ Data Quality



SARIMAX

8%

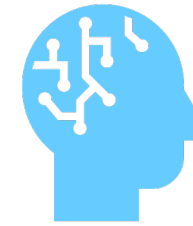
■ Visualizations are key

Future Work



User Study

After building an MVP, we'd like to observe how Z-POP influences the behavior of its users and if we are able to prevent overruns.



Models

We want to test more expressive, but still interpretable models (e.g., TFT) and incorporate more data (e.g., user-specific time bookings, holidays, expenses).

Detecting Project Overruns with Time Series Models

Thanks!