

Estimation of Application Maintenance by Means of Machine Learning

ICEAA Workshop 2022

Eric van der Vliet, Direct Estimation CoE

Kimmo Kettula, Senior Data Scientist CGI Advanced Analytics Solutions



Agenda

- Application Maintenance Estimation Initiative
- ScienceOps Approach
- Estimation Environment inputs
- ML Model and results
- Next steps
- Estimation Environment reporting

CGI at a glance

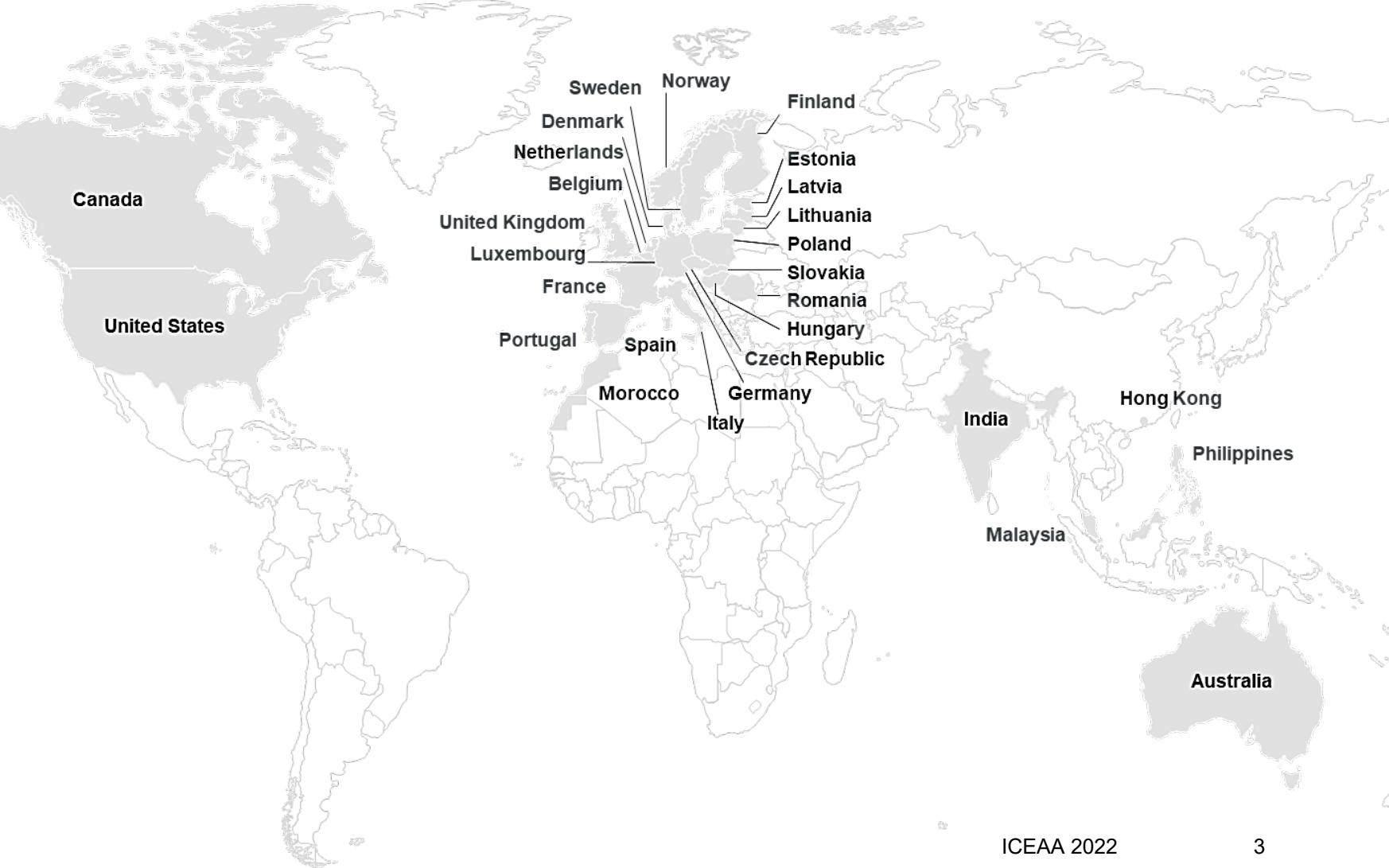
Founded in 1976
45 years of excellence

CA\$12.1 billion revenue

82,000 consultants

400 locations in 40 countries

5,500 clients benefiting from end-to-end services



Helping clients
become digital through
**end-to-end
services**

Including Application Maintenance

CGI



Application Maintenance Initiative

Objectives Application Maintenance initiative



Estimation Improvement

Improved estimation

- More accurate estimation models / benchmark facilities based on historical data
- Estimates based on data of comparable engagements
- Estimates with a realistic impact of efficiency levers over time (multiple years)

Improved data analysis

- Exploiting data to make data driven decisions (*#Tickets, Quality, Finance*)
- Data analysis to improve the application maintenance solution engineering
- More impact of efficiency levers over time (*e.g. automation, robotics, ...*)

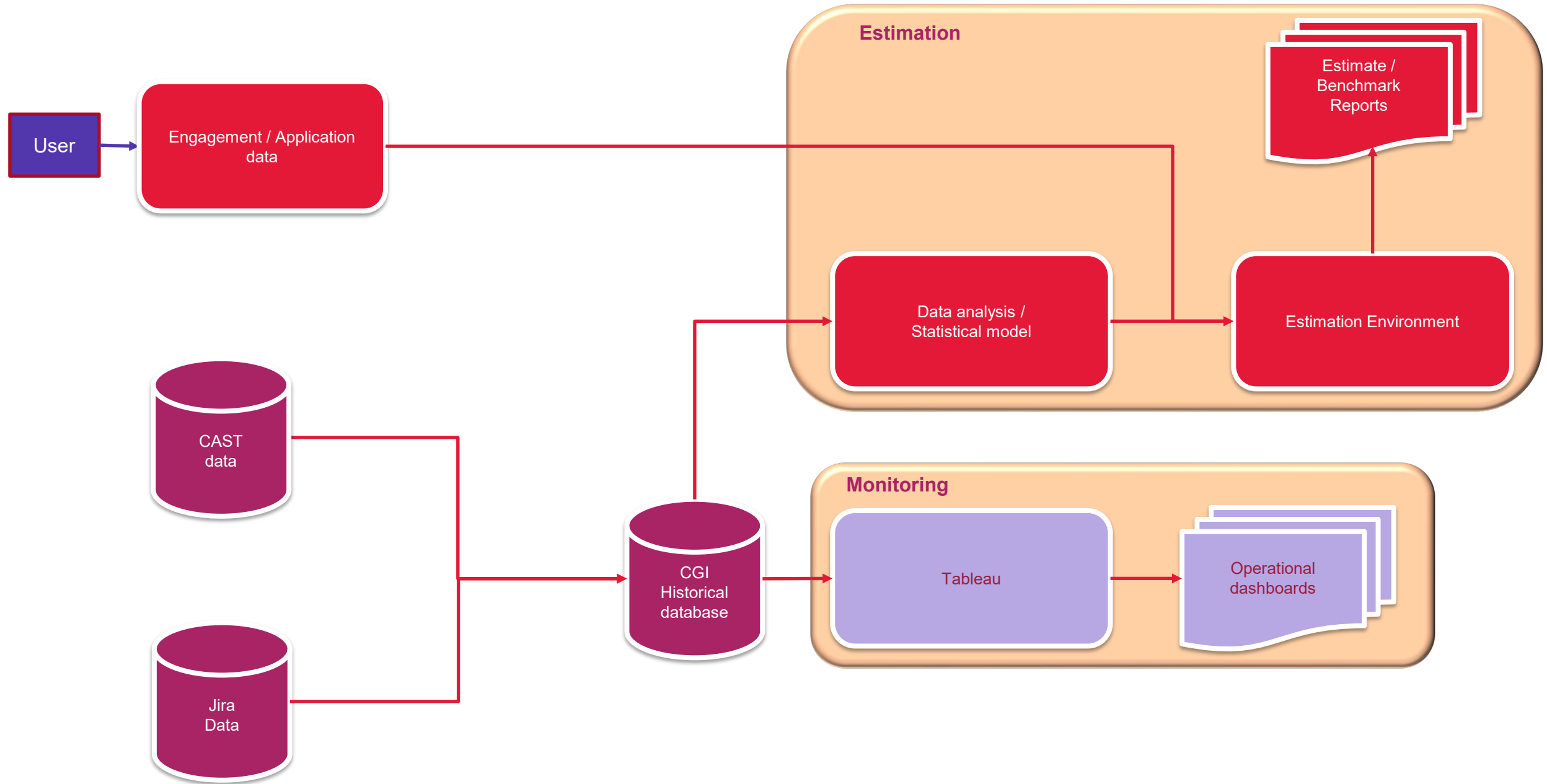
Operational benefits

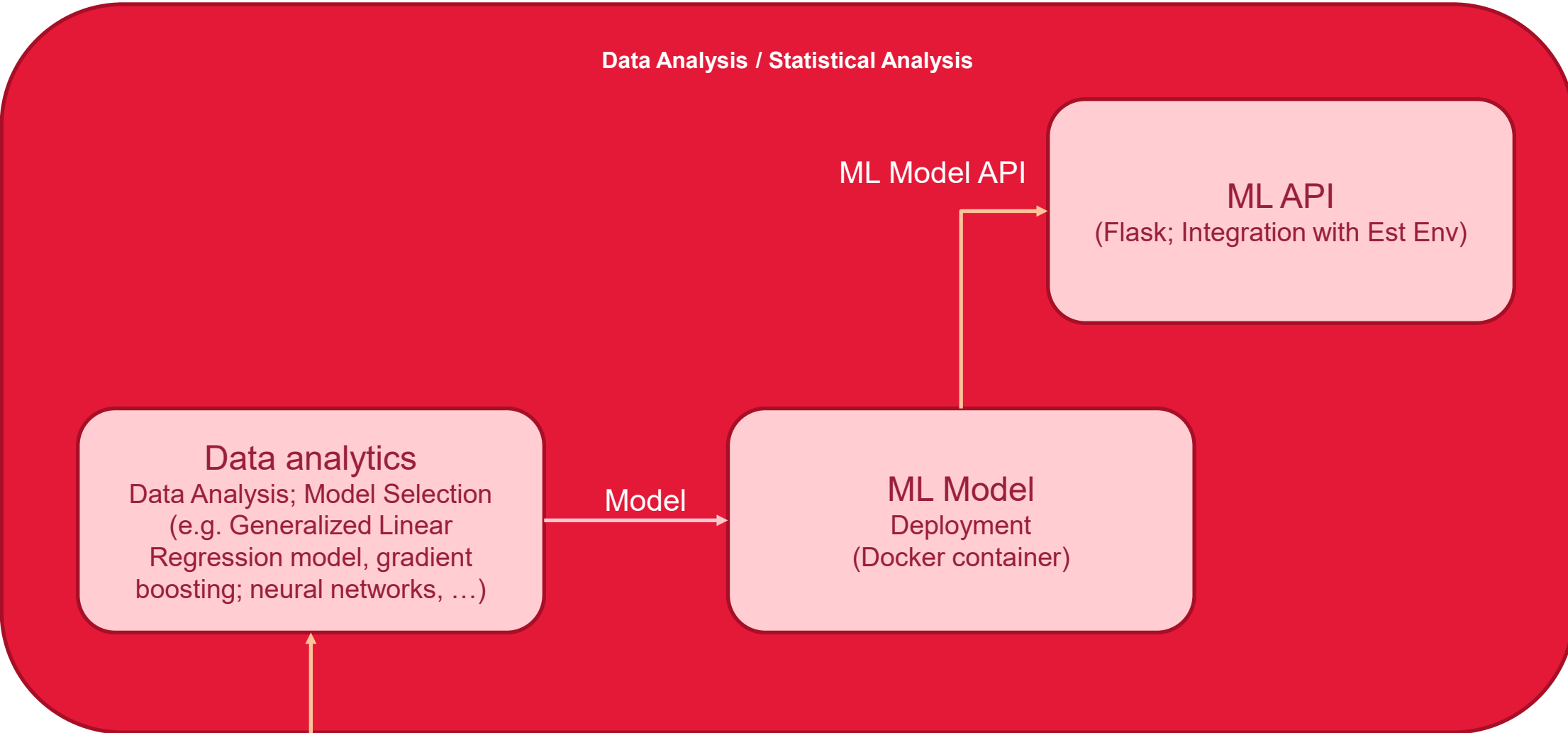
Improved monitoring

- Improved monitor of delivery efficiency and actual ticket costs
- Holistic view of the data; effort and cost
- Identify further efficiency improvements that could be achieved

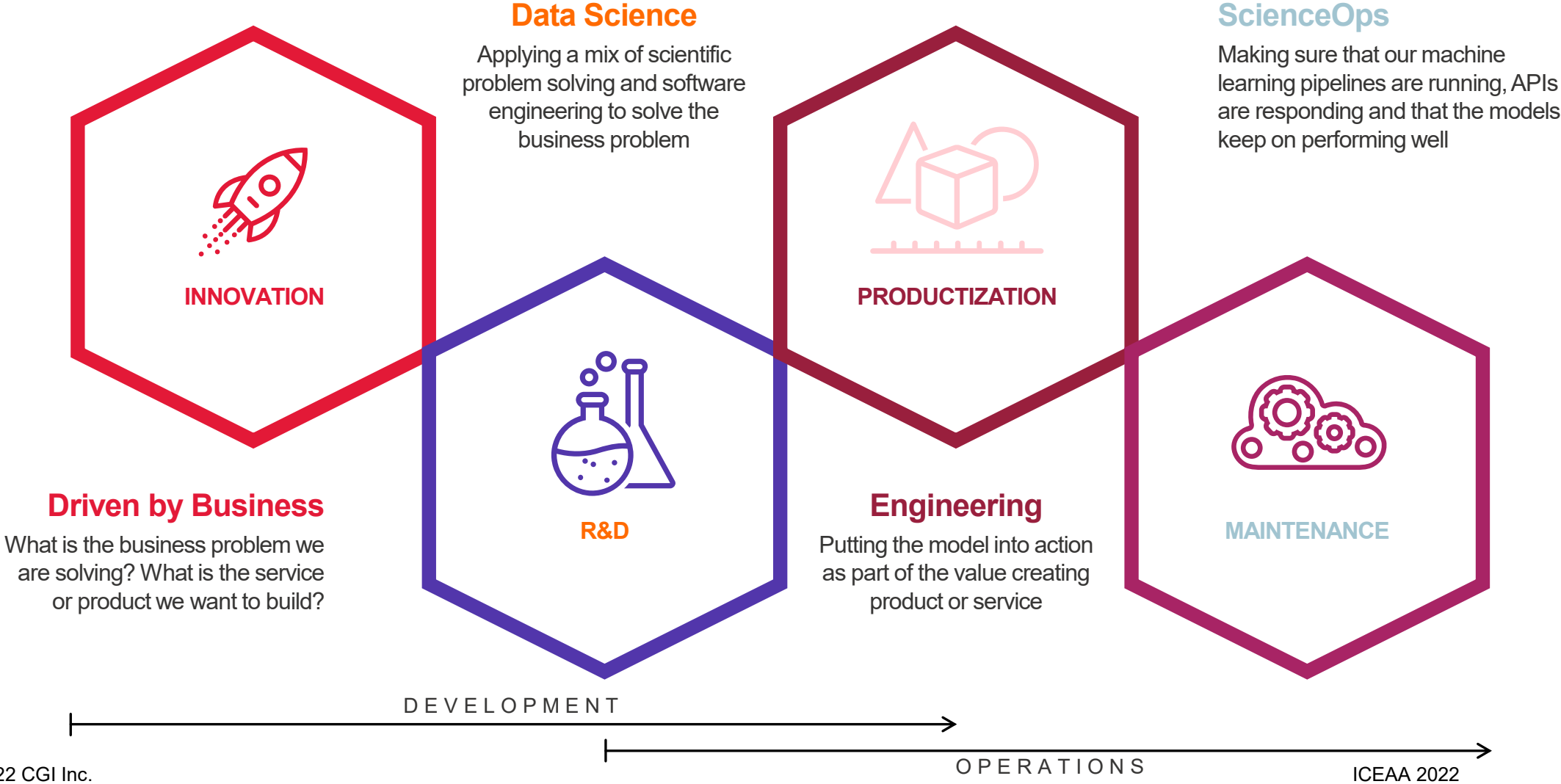
Improved data registration

- Consistent data registration (e.g. effort per ticket)
- Extension of data registration (e.g. technology, complexity)
- More pre-formatted and mandated fields in registration systems





Machine Learning solution development



Estimation Environment input

Test UAT Releases

Details Upload Assigned Apps

Upload Data

AM Estimation model File Name.xls ✓ uploaded
 40 KB | 10/12/2021 10:56 AM
 Delete | Replace

*Add the right worksheet to your excel before you upload. [Click here to download the template](#)

+ Add New Row Edit Delete Download Excel

Application Name	ApplicationID	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Select All
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>
Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	Lorem ipsum	<input type="checkbox"/>

Save Cancel Create Solution

Loading of app details
 Ability to add current situation
 to be able to derive efficiency levers

1 Select Estimation Model

Estimation model
 AM Estimation Model ⓘ ⬇ Skip to Estimation Breakdown

Application Portfolio

No. of Applications	Description	No. of Tickets	Technology	Domain	Complexity	Skill Level
1234567	Lorem ipsum dolor sit amet, consectetur adipiscing	123456789	Lorem ipsum dolor	Industry	Average	Average

Activate

Input parameters
 Selection of benchmark data

Hours spent per ticket : Profile 1

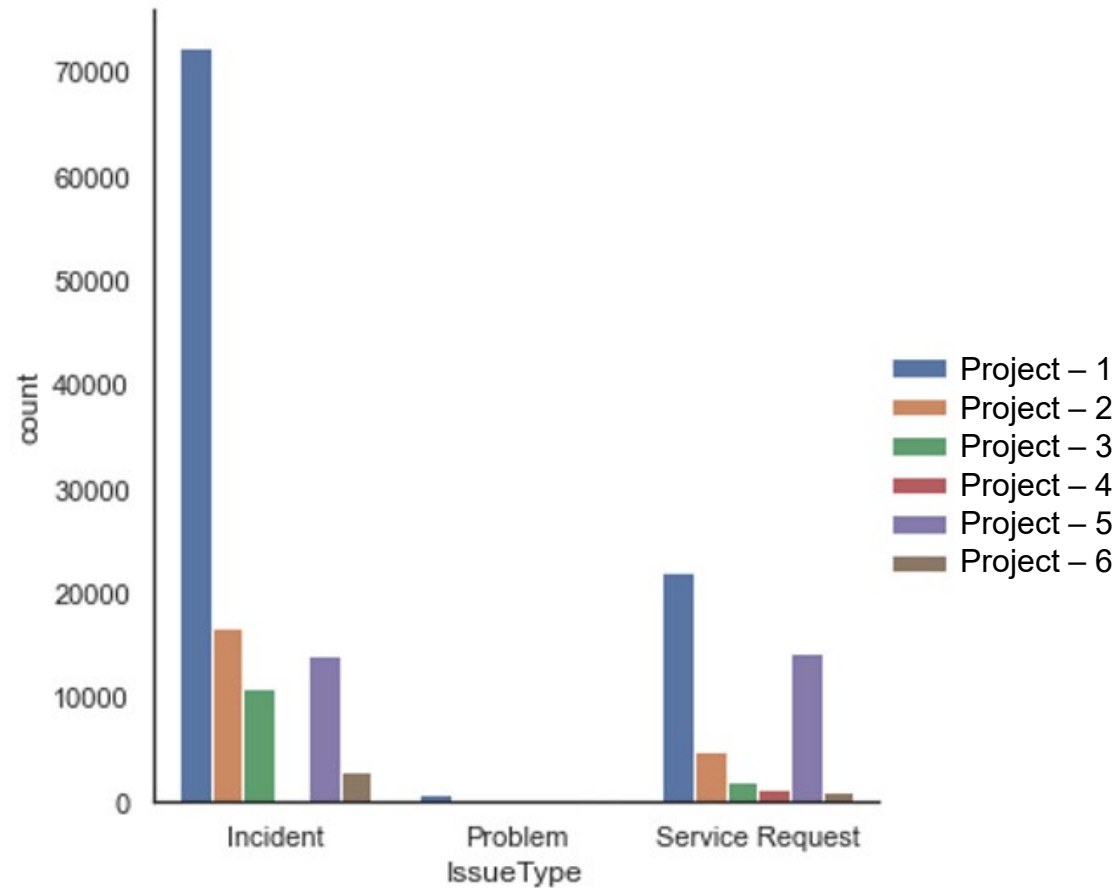
	Priority 1	Priority 2	Priority 3	Priority 4	Average
Problem	1.10	1.05	0.95	0.88	1.00
Incident	2.30	2.15	2.10	1.90	2.11
Service Request	2.30	2.15	2.10	1.90	2.11
Total	1.90	1.78	1.72	1.56	1.41

% of ticket based effort

Service Monitoring	10
Service Operation	2
Service Management	5
Total	17

Estimation results
 From historical data;
 Determined based on selection
 (ML Model)

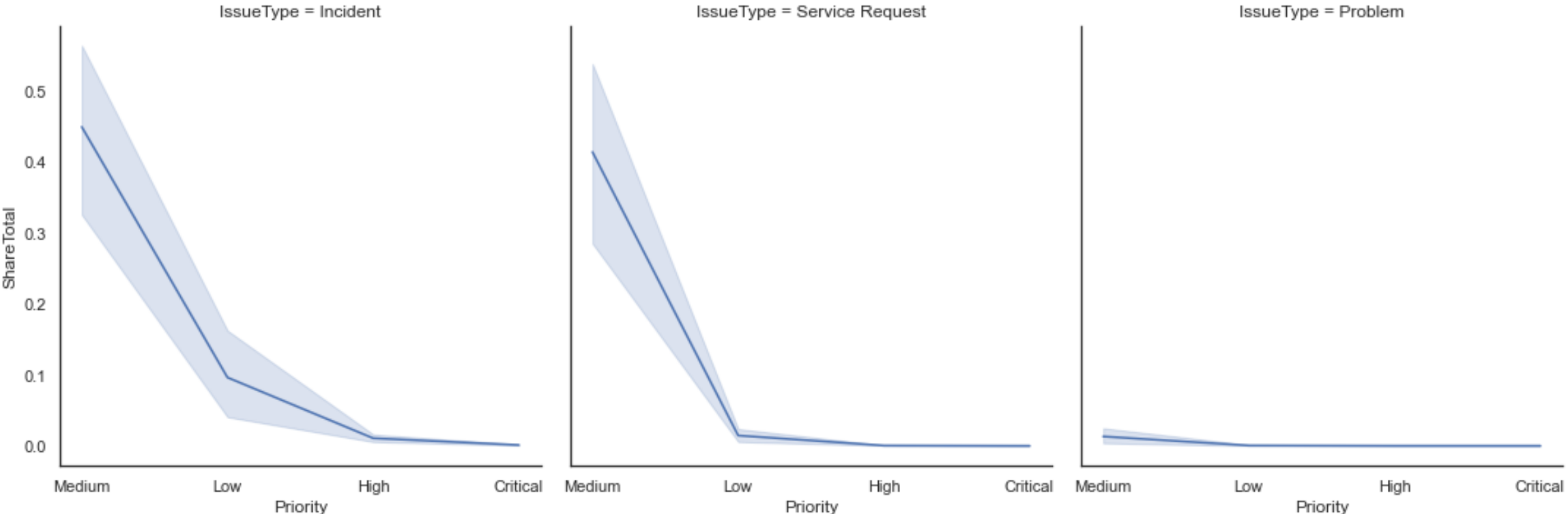
Understanding the data - 1



Observations

- **Mainly Incident and Service Request data**
- Limited amount of Problems
- Large variation in numbers depending on the project
- Project characteristics required to determine the differences

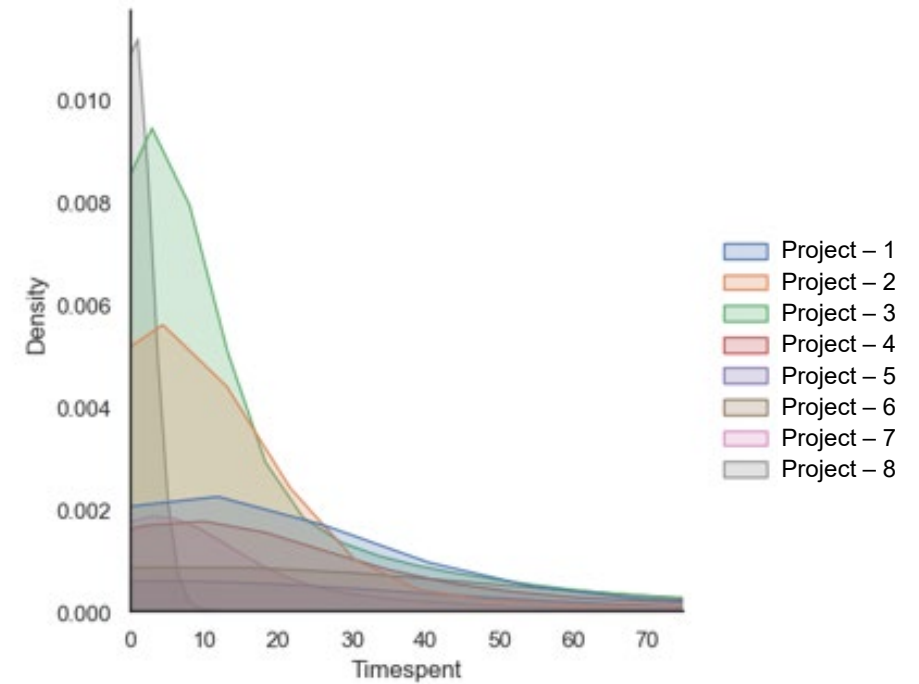
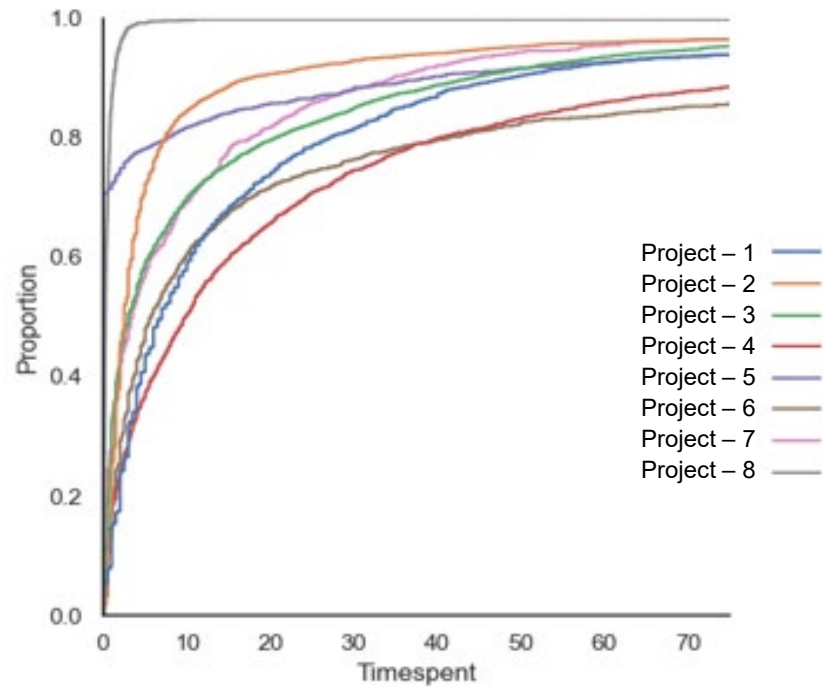
Understanding the data - 2



• Ticket distribution with 90% confidence intervals

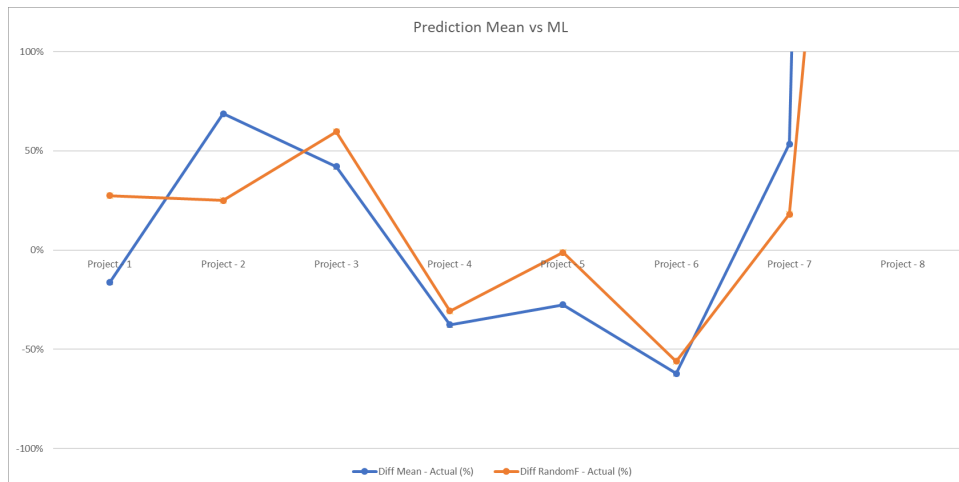
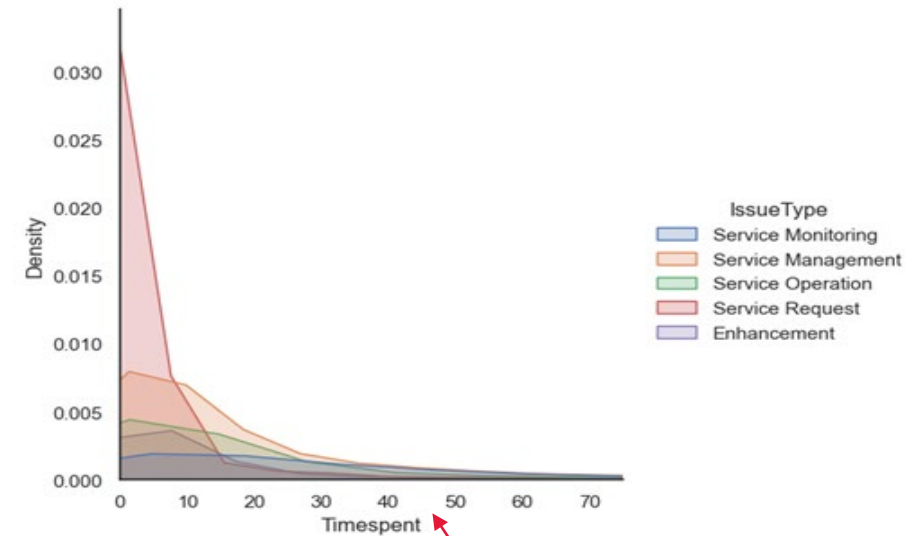
Understanding the data - 3

- A small amount of data has the highest effort
- This long tail in the effort will influence averages
- Another analysis is required to come up with an effort prediction



Understanding the data - 4

IssueType	50th percentile	80th percentile	95th percentile
Enhancement	2.5	10.4	33
Service Management	3.6	23	41
Service Monitoring	11.0	33.2	83
Service Operation	2.1	16.8	56
Service Request	0.6	2.2	17

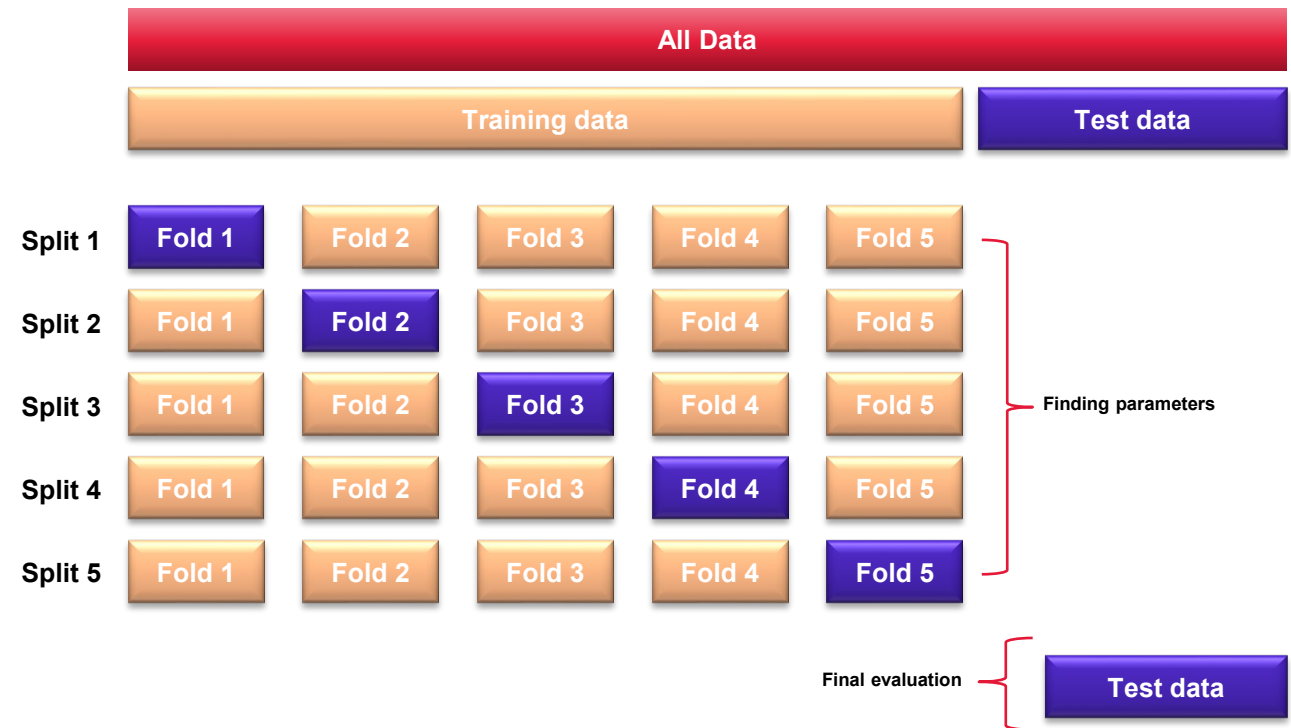


Conclusions

- **Median values are influenced by a long tail in the actual issue type data**
- **Estimate based on mean values can deviate 50% (or more) from actual data**
- Statistical model is strongly depending on the project selection
- 80th percentile of historical data provides the best prediction
- Further refinement of model requires more detailed data (e.g. technology; domain)
- Without additional details, historical data must be selected carefully

Model and modelling approach

- The best performing algorithm based on the first set of data is: Tweedie regression (Generalized Linear Regression model)
 - Works best out of a number of algorithms
 - Useful in cases where there is a spike at or near zero and a long tail of positive values
- Model training and parameter tuning is performed using k-folded cross-validation
- Test of performance metrics on 12 months of data
- The best fitting tuned model is trained using all available data



Reporting - 1

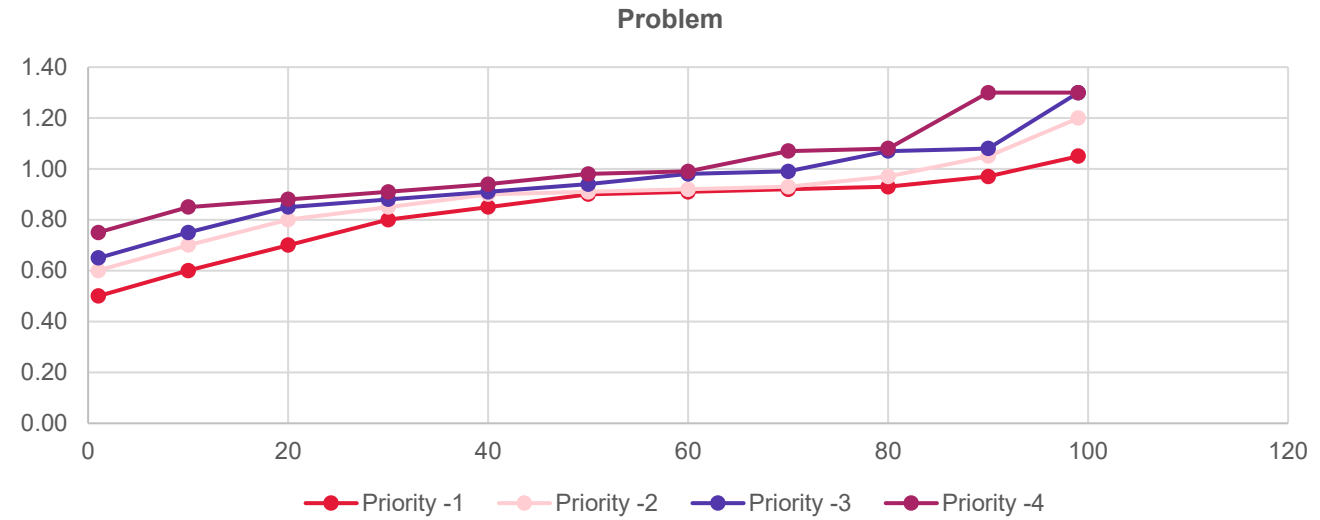
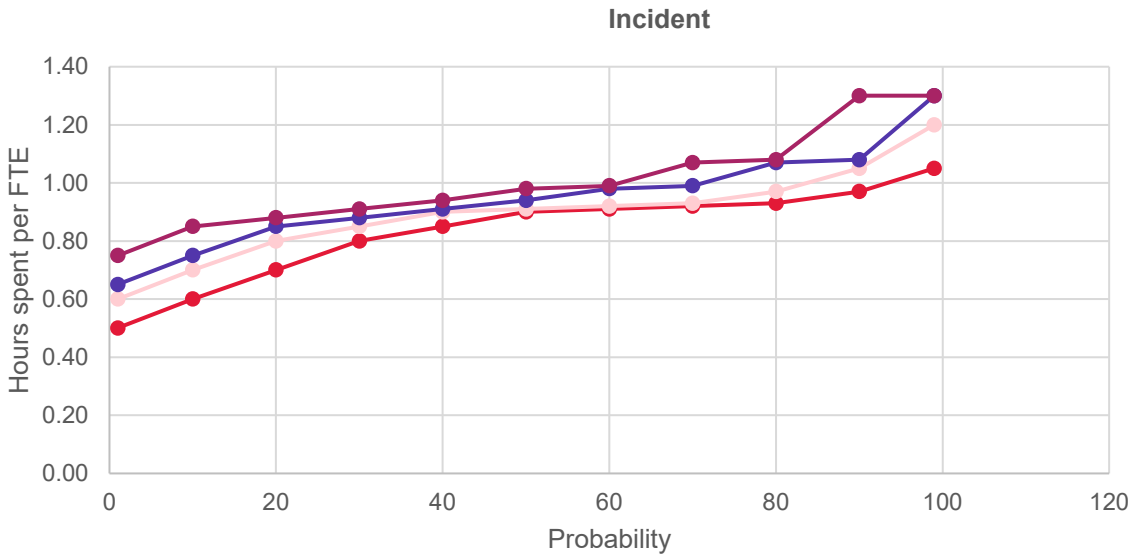
Total Ticket Effort (Direct) : 26535	Total Non Ticket Effort (Direct) : 3114	Total Non Ticket Effort (In Direct) : 1857
---	--	---

	Cost Type	Priority 1	Priority 2	Priority 3	Priority 4	Total
Problem	Ticket Direct	825	1103	1140	1320	4388
Incident	Ticket Direct	2415	2580	2835	2850	10680
Service Request	Ticket Direct	3450	2903	2835	2280	11468
Enhancement	Non Ticket - Direct					460
Service Monitoring	Non Ticket - Direct					2654
Service Operation	Non Ticket - In Direct					531
Service Management	Non Ticket - In Direct					1327

Total Ticket Effort (Direct) : 44254	Total Non Ticket Effort (Direct) : 10115	Total Non Ticket Effort (In Direct) : 3096
---	---	---

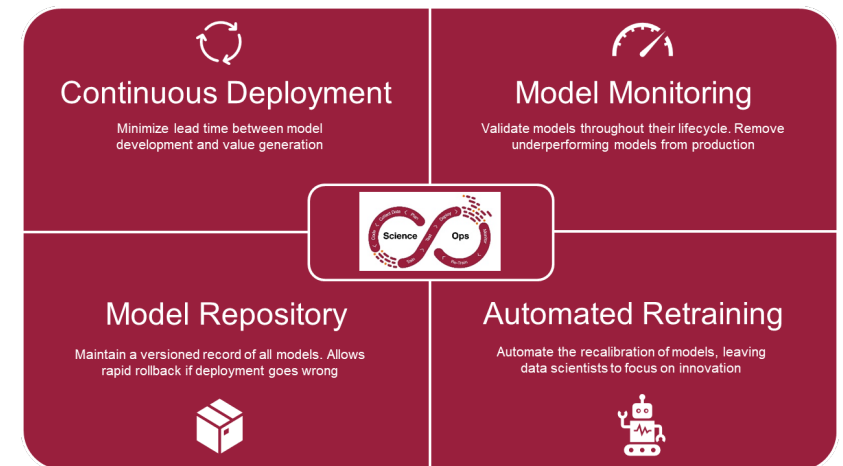
	Cost Type	L1	L2	L3	Total
Problem	Ticket Direct	-	2,193.75	5,118.75	7,313
Incident	Ticket Direct	-	17,800.00	-	17,800
Service Request	Ticket Direct	-	-	19,112.50	19,113
Enhancement	Non Ticket - Direct	-	-	-	5,692
Service Monitoring	Non Ticket - Direct	-	-	-	4,423
Service Operation	Non Ticket - In	-	-	-	885
Service Management	Non Ticket - In	-	-	-	2,211
Total		-	19,993.75	24,231.25	57,435.25

Reporting - 2



Next steps

- Refining the model taking into account more application characteristics
- Refining the model taking into account efficiency improvements based on efficiency levers
- Further improvement of the source data (Jira, CAST, Sonar, ...)
- Continuous improvement of the statistical model and deployment
 - Gradient Boosting
 - Machine Learning



Questions?



Thank you

Eric van der Vliet

E-mail: Eric van der Vliet

Tel: +31612872259

Linked-in: ericvdvliet

Twitter: @ericvdvliet

The CGI logo is displayed in a bold, red, sans-serif font.