5G rollout

Gemma Vall Llosera, PhD
Researcher at LTH Elektro- och informationsteknik
Principal Q engineer at Ericsson

Agenda

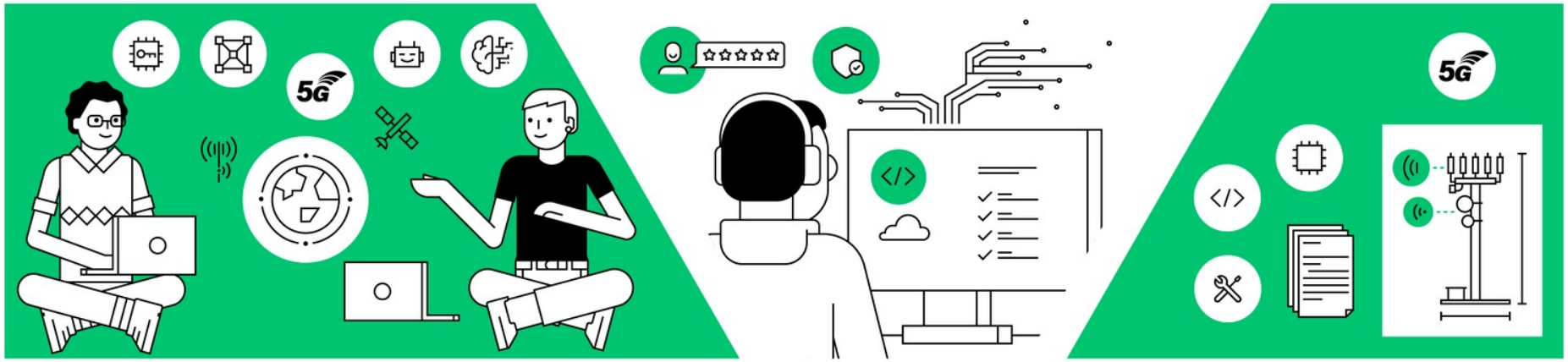
- 5G end-to-end storyboard
- 5G rollout process
- a practical example of a data science use case

5G end-to-end storyboard

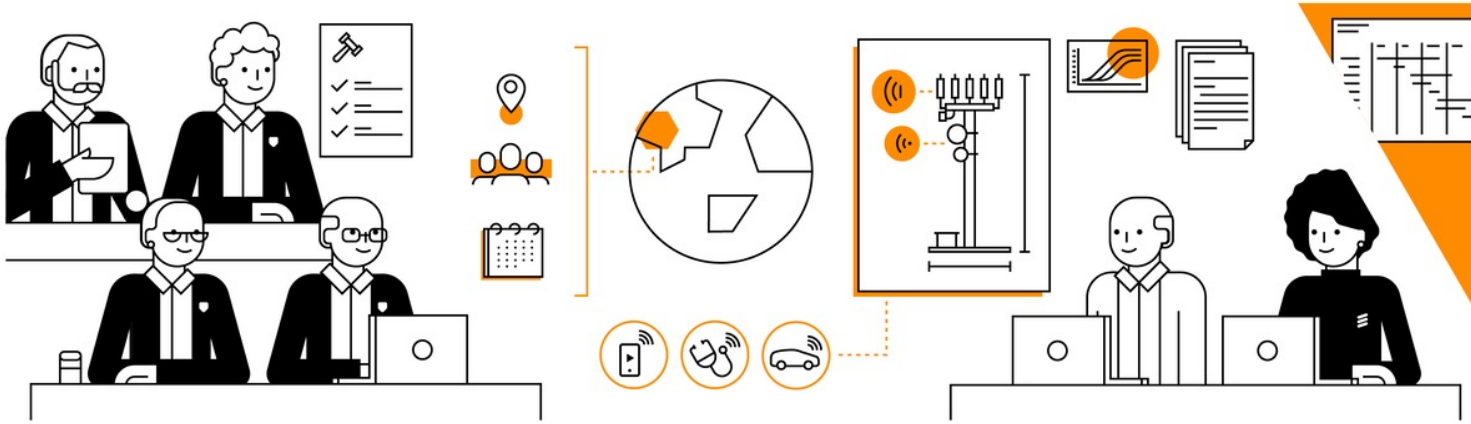
Gemma Vall Llosera Automation & AI

2023

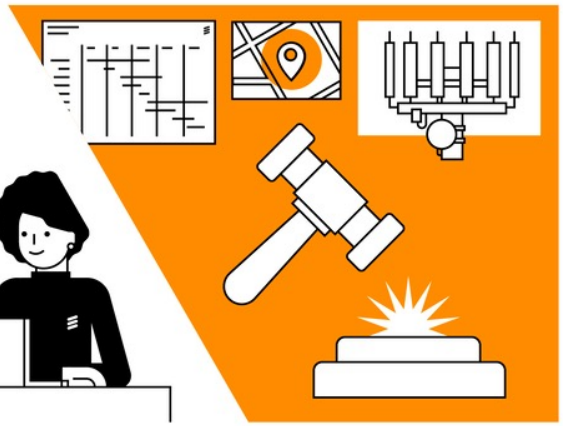
01. Standardization and product development



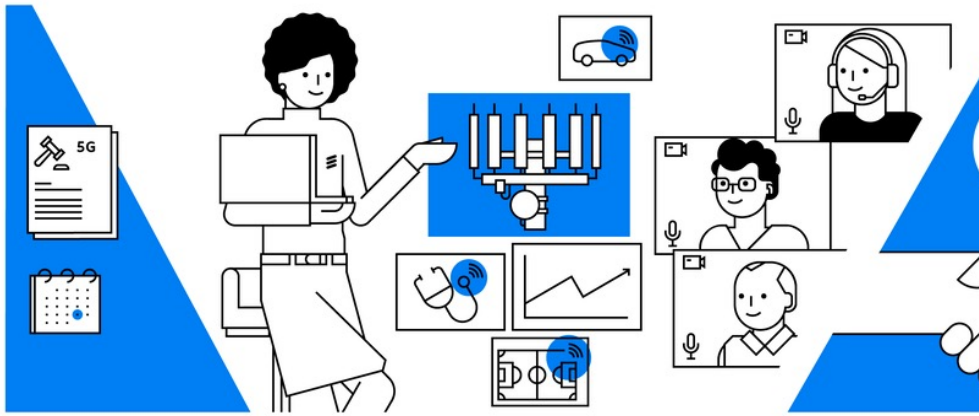
02. Government & Telecoms: 5G Spectrum auction



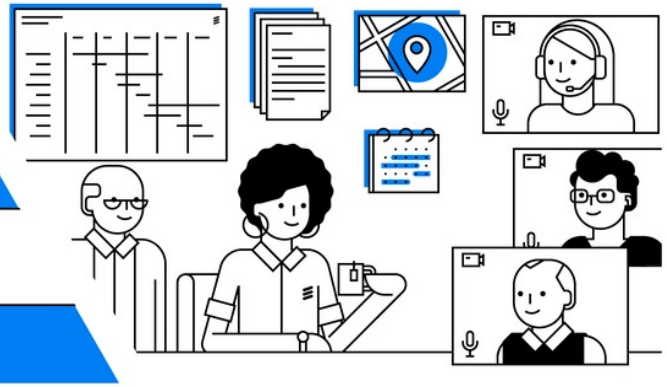
03. Spectrum licenses awarded to telecom operators



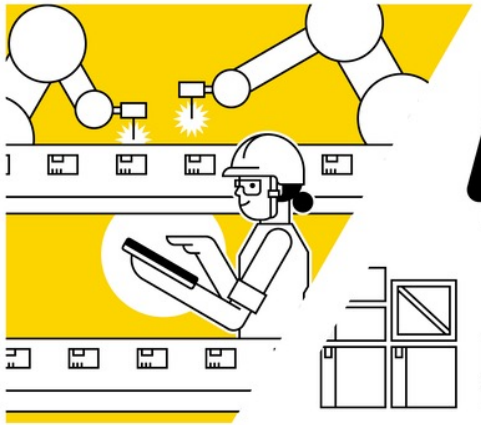
04. Ericsson + Operator: Sales process



05. Completing order details



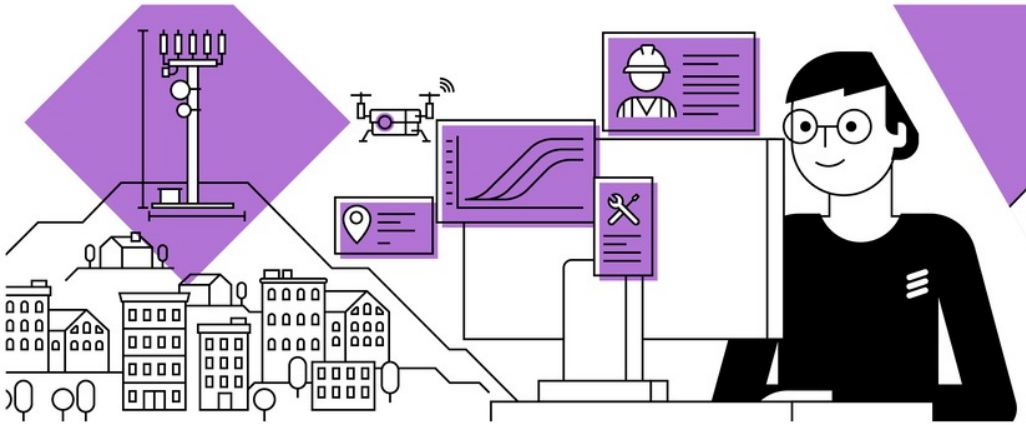
06. Factory: Manufacturing equipment



07. Warehouse: Organizing and storing equipment

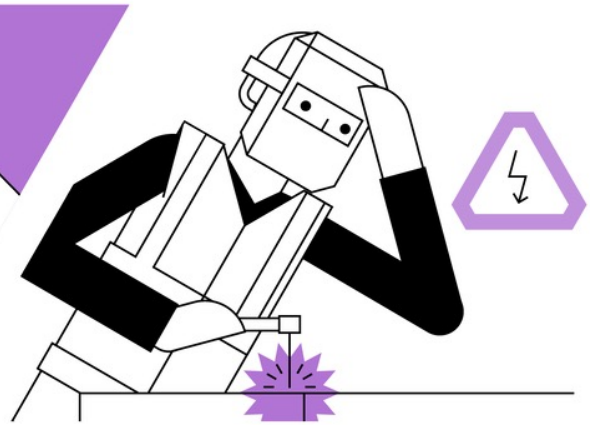
08. Freight shipping

09. Site survey: Specifying construction detail



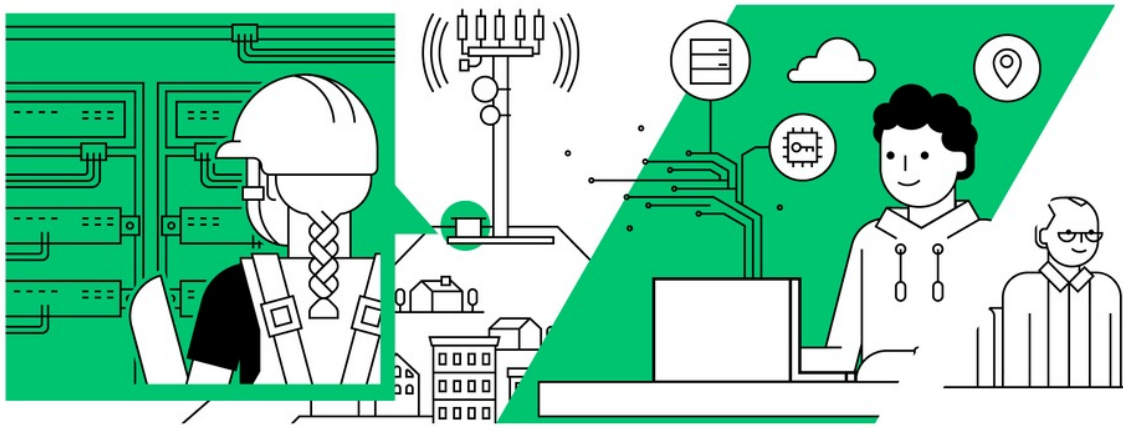
10. Field: Preparing terrain

11. Field: Assembling equipment

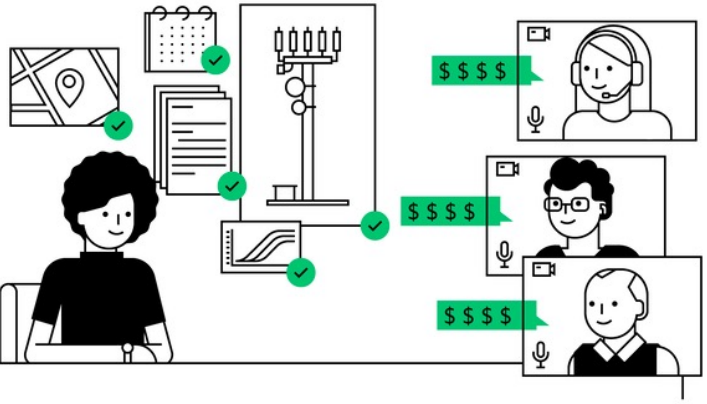


12. Constructing site

13. Connecting site to network

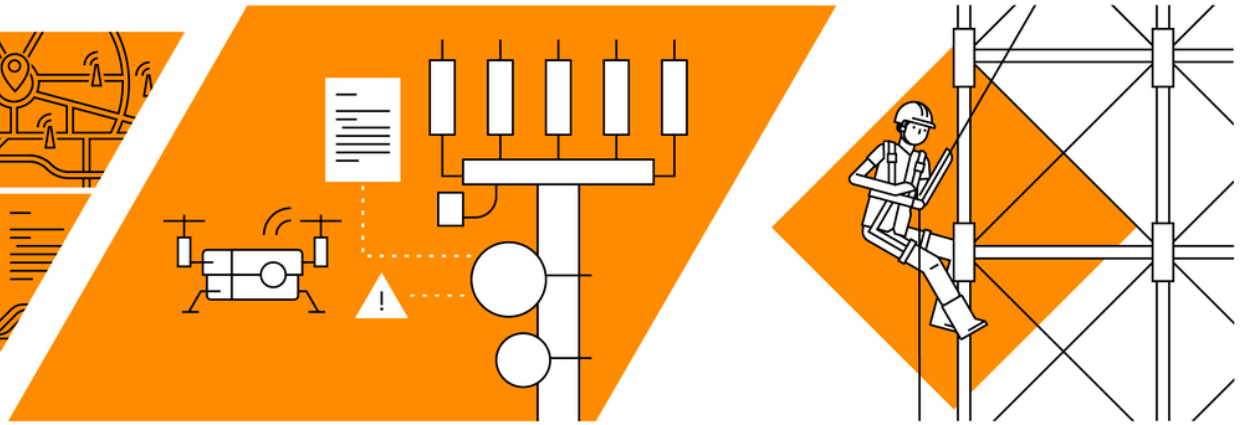


14. Site acceptance and payment



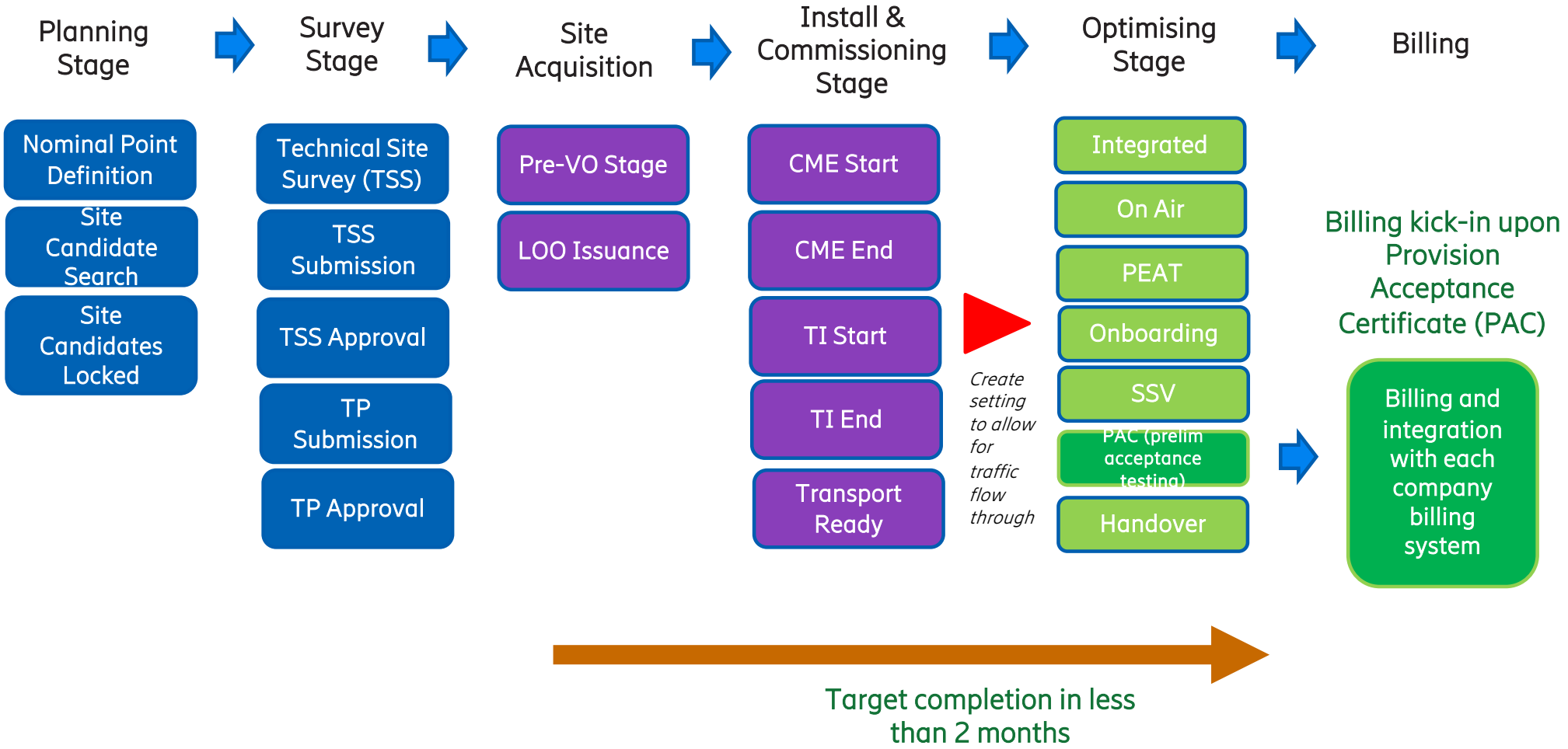
15. Remote support

16. Field maintenance

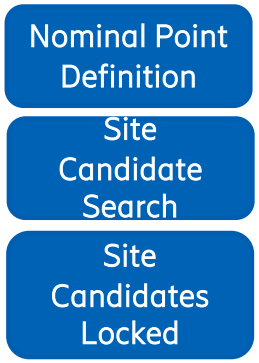


5G rollout

5G rollout milestones

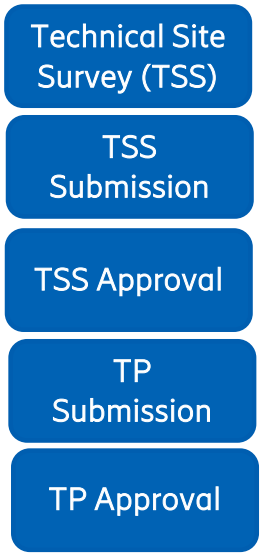


Planning stage



- After network dimensioning comes the planning stage
- Planning stage: 3 lat/long are chosen as candidates to deploy the sites
- Nominal point definition: network dimensioning based on throughput, latency, availability, etc
- Site candidate search: based on the output of the network dimensioning, candidate locations need to be explored
- Site candidates locked: after exploration, sites are locked
- Typical duration: 7 days

Survey stage



- Survey stage: from visiting the site to approval of the technical plan by the customer
- Technical site survey (TSS): physical visit to the sites to see floorspace availability, power availability, previous installation present, building or mountain, distance to closest road, paved road? fibre optical present? can we dig? Out of the 3 candidates, which one is available?

- TSS submission: submission of the report to the customer- this is my plan for the rollout
- TSS Approval: Customer approves the report, usually on 3 sites
- Technical plan (TP) submission: at ths stage only one site is chosen
- TP Approval: approval of the plan by the customer

- Typical duration: 26 days

Site acquisition

Pre-VO Stage

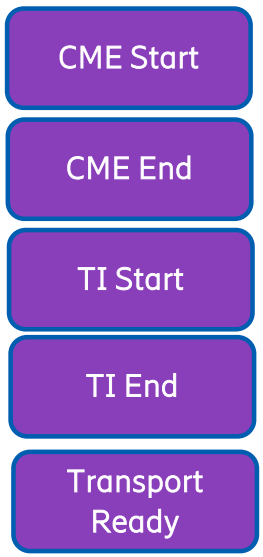
LOO Issuance

- During the site acquisition we need to do research, evaluate, and negotiate leases for locations that can host telecommunications equipment such as wireless transmission towers. This is done during the pre-visual object stage.

- Letter-of-offer (LOO) Issuance: letter of offer to the land owner /landlord of the building with the details of the contract

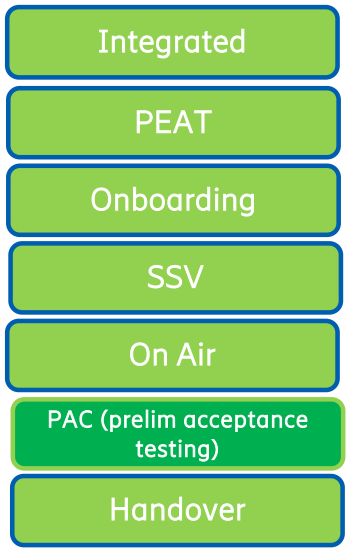
Once the site is bought the install & commissioning stage starts

Install and comissioning stage



- This stage goes from the start of the CME work until the transport is ready.
- Civil, mechanical and electrical work (CME) includes civil works on ground and on air, ex. structural stengthening, 3-phaese power, clear any debris, ground need to be compacted with gravel,
- Technical installation (TI) includes the installation of power and building up the radios – digital unit, remote units.
- Transport ready: when the transport network is ready. by transport we mean optical fibre and/or microwave.
- typical duration CME start to CME acceptance between 37 or 67 days depending on if it's collocated or new site.
- typical duration from TI end to transport ready is 8 days for sites with fibre backhaul, 12 days for sites with microwave backhaul.
- note that if fibre is not present and it needs to be deployed, this work already starts after the TSS (technical site survery) and it takes around 90 days.

Optimisation stage



- the optimisation stage starts with the integration finished and concludes with the handover to the customer
- Integration: sites are installed, core network is connected, i.e. fibre and MW working
- Prelimindary equipment acceptance test (PEAT):
- Onboarding of the actual users
- Single site verification (SSV): verification step
- Provision Acceptance Certificate (PAC): preliminary acceptance testing
- On-air: in this phase a few phone calls are made

- Handover: fully accepted by the customer
- Typical duration: this varies a little bit as in here we have several different workflows. It could be between 52 and 79 days

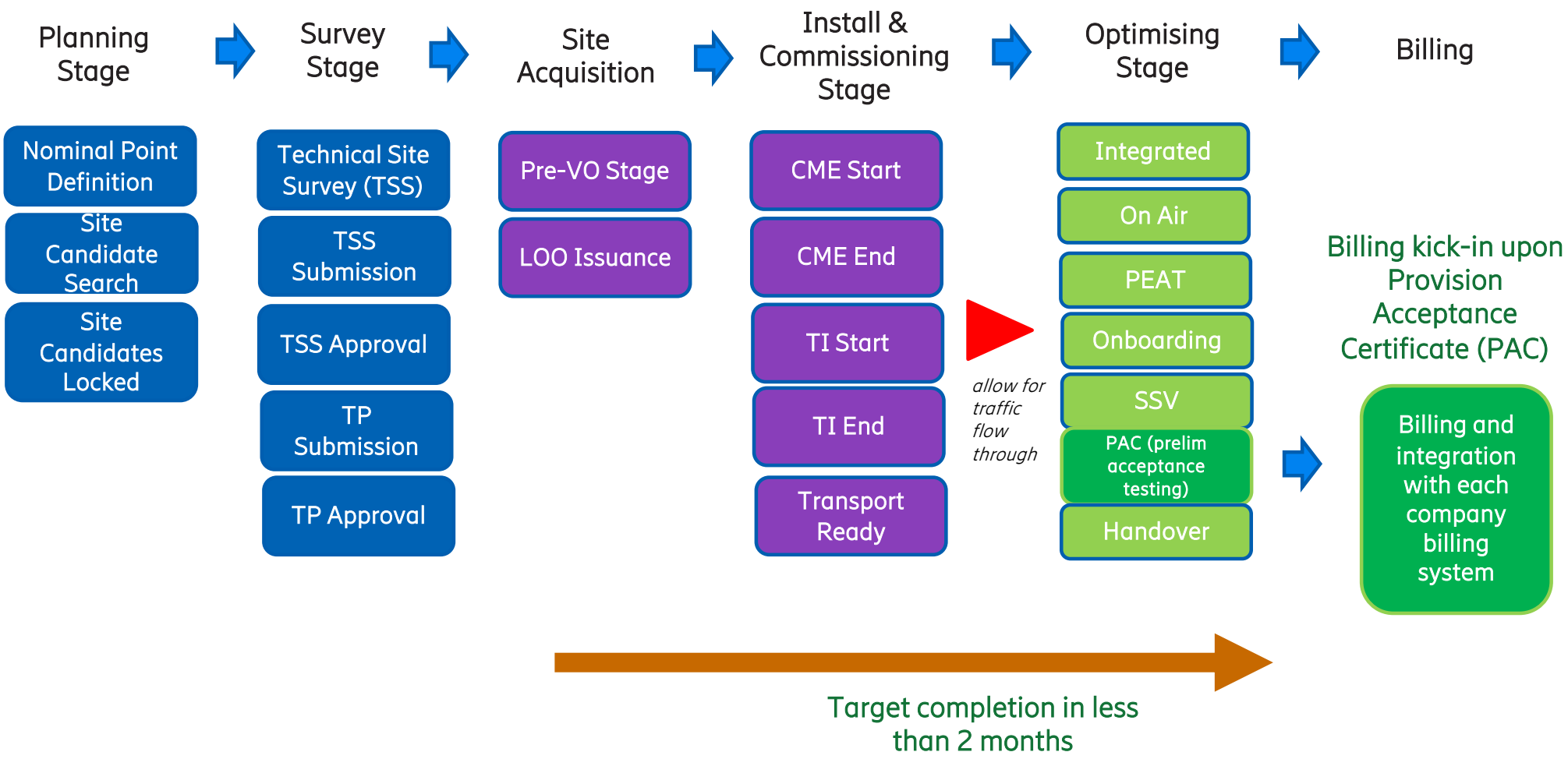
Billing

Billing kick-in upon Provision Acceptance Certificate (PAC)

Billing and integration with each company billing system

- In this stage the billing gets integrated

5G rollout milestones

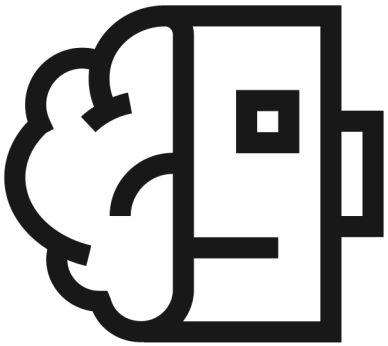


A practical example of a data science use case

5G rollout milestone prediction

- We have seen a 5G rollout goes through many milestones

- What if given the date when the TSS is approved, site characteristics, radios to be deployed, etc we can predict when the site will be installed? integrated and on-air?
- what if we could classify if the site will be finished in more or less than 120 days?
- this would allow us to send the right people to the right site depending on prioritisation of sites

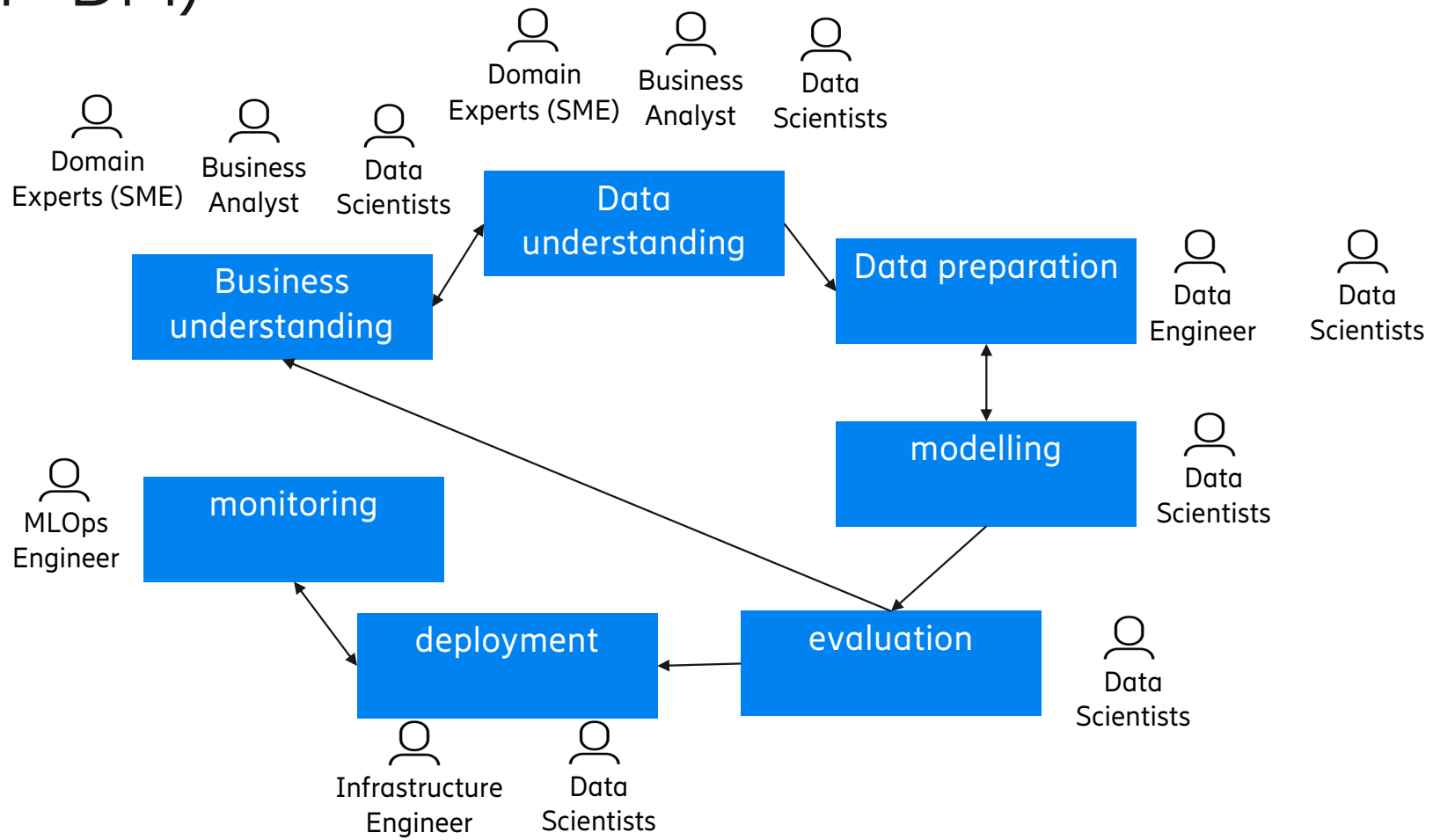


Data science holy grail

- Ask the relevant question and using the domain knowledge to capture the right data

- Do NOT collect too much data
- You need to be a good filter

Cross-industry standard process for data mining (CRISP-DM)



Use case matrix

Purpose Goal of the project	**Value Created** Key benefits of the projects	**Initial Hypothesis** What do we hope to find
Data What input data is available? What is he data quality?	**Activities** Steps of the project	**Possible risks** Issues that can impact the project

In order to fill in the use case matrix

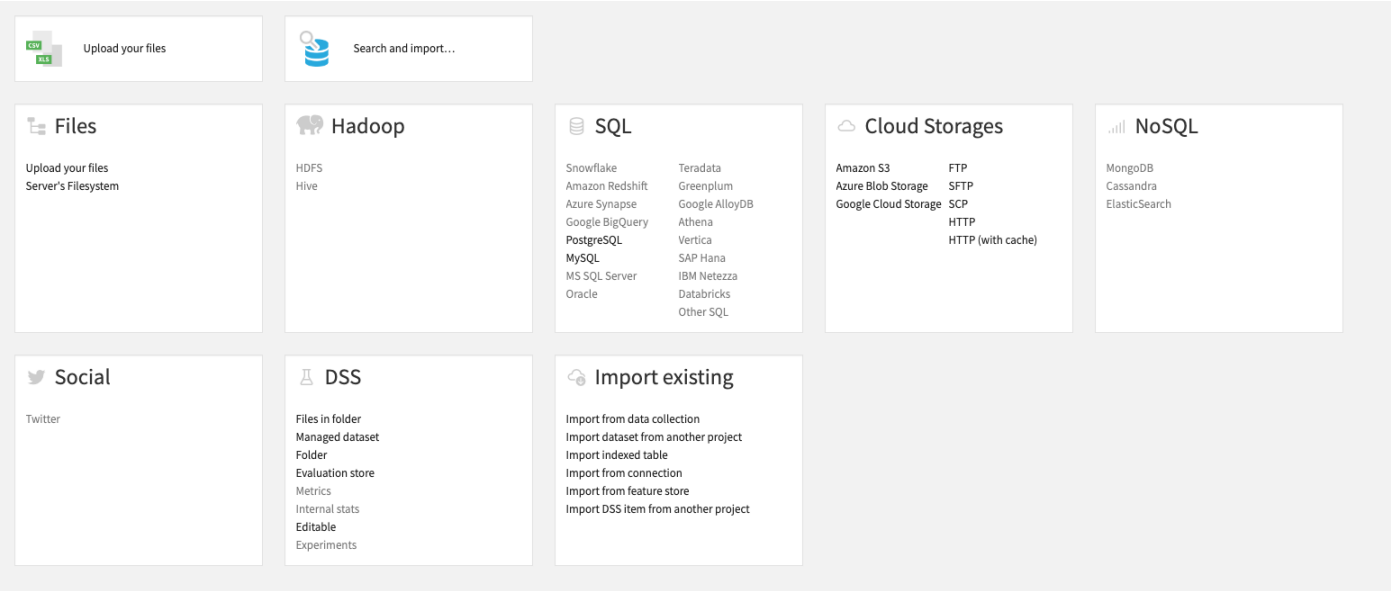
- Business understanding
- meet with stakeholders and subject matter experts
- discuss the pain points
- translate the use case into a data science use case
- define who will use the results and how they need to be visualised
- define success criteria
- produce a project plan, budget and team that will deliver the use case

Use case matrix

Purpose Goal of the project	**Value Created** Key benefits of the projects	**Initial Hypothesis** What do we hope to find
Develop prediction and classification models for installation, integration and on-air milestones	• Better resource allocation • better customer acceptance • less delays • automated reporting	• Visualise site progress against plan • Forecast actual plan to prevent delays • Stie reprioritization in case of forecasted delay
Data What input data is available? What is he data quality?	**Activities** Steps of the project	**Possible risks** Issues that can impact the project
6000 rows 1700 cols with data from planning, survey, site acq, install & comissioning, optimising stages	Planning, data ingestion, transformation, modelling, deployment, test and validate, publish data on dashboards	Geographical admin data is not available for polygon visuals Data collected is not representative of geography to predict

data ingestion

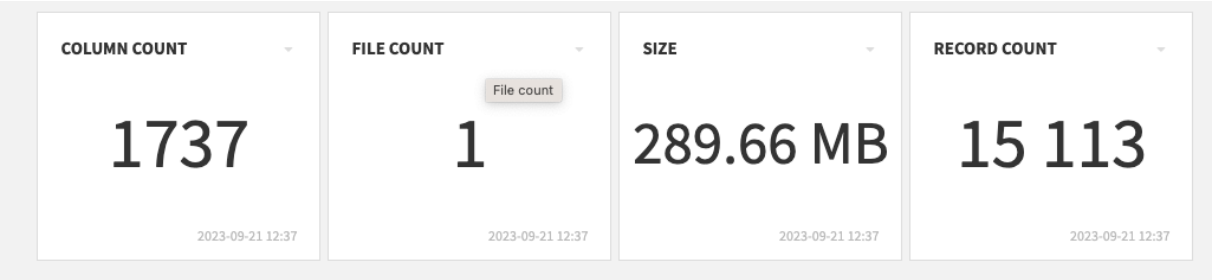
- Data resides in a specific country and can't leave the country
- Data is stored in a database without API
- Is this tabular data?
- is this parket data?
- is the network port to that database open?

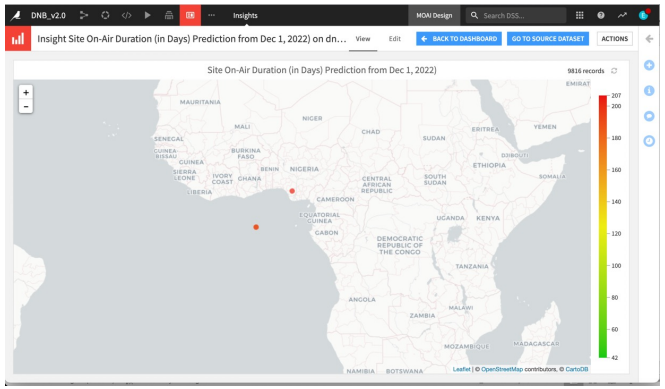


Data and Data Quality

- Data Source
  - DPM Azure Data blob

- Data Quality
  - Timeliness : High
  - Validity: High
  - Consistency: Low
  - Integrity: Low
  - Completeness: Medium
  - Volume of data: Medium
  - (15k rows and 1737 columns, but many empty)





Data Quality- volume of data

– we could score more sites going back to just TSS_approval_actual

- Integration data analysis
  - Start with 14059 rows
  - 7628 rows after removing Site_Status = inactive
  - 6893 rows after removing empty values in TSS_Actual_Date
  - Then we split 3824 rows for unseen Integration_duration and 3184 rows for seen (we could score 3824 sites)
  - In reality we score 331 rows (8.33%) because the num of rows in Installation duration = num of rows in Installation_Actual_end and it's only 8.33% full

- Let's remove almost empty features: Transmission_req_actual, Fiber_acceptance_actual, RAN_CME_actual_end, installation_actual_start, installation_actual_end and installation_duration
- we still score 331 rows
- because:

1. Fibre _ready_last_point_actual has 554/1662 rows with invalid data `2000-01-01`
2. CME_Actual_end has 886 rows
3. CME_actual_start has 1086 rows of data
4. TP_approval_actual has 3266 rows of data

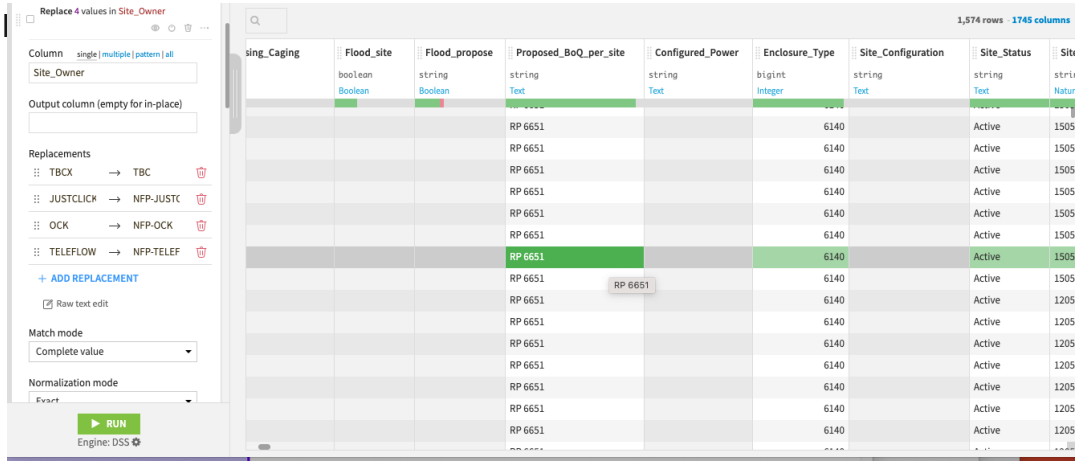
The scoring (prediction) only works for overlapping rows (all columns have a value in the same row)

1,2,3,4 only overlap in 331 rows...

- Integration duration is calculated as = TSS_Actual_Data – Integration_Actual

data preparation

- Remove useless columns
- Find and replace : 30m, 30M, 30, thirty meters, 30 |
- create geopoints
- compute time difference
- Remove rows with empty values
- Filter rows where site is not active
- Replace "" by 0
- Fill empty cells with ...
- Replace 2 values in building height
- parse data
- ...



Modelling

- Iterate with different input features
- Iterate with different algorithms
- hyperparameter tuning
- model evaluation and comparison (metrics, exectution time...)

Algorithm

Algorithm details

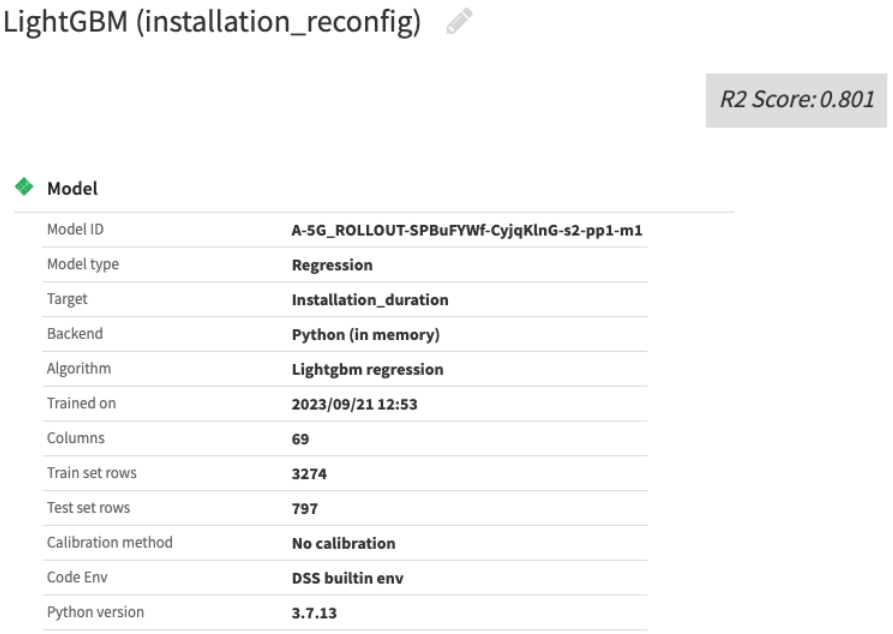
Algorithm	**LightGBM**
Booster	**gbdt**
Actual number of trees	**54**
Maximum number of leaves	**31**
Learning rate	**0.1**
Alpha (L1 regularization)	**0**
Lambda (L2 regularization)	**0**
Minimal gain to perform a split on a leaf	**0**
Min sum of instance weight in a child	**0.001**
Subsample ratio of the training instance	**1**
Columns subsample ratio for trees	**0.7**

Training data

Rows (before preprocessing)	**3274**	Rows (after preprocessing)	**3274**
Columns (before preprocessing)	**69**	Columns (after preprocessing)	**966**
Matrix type	**dense**		
Estimated memory usage	**24.13 MB**		

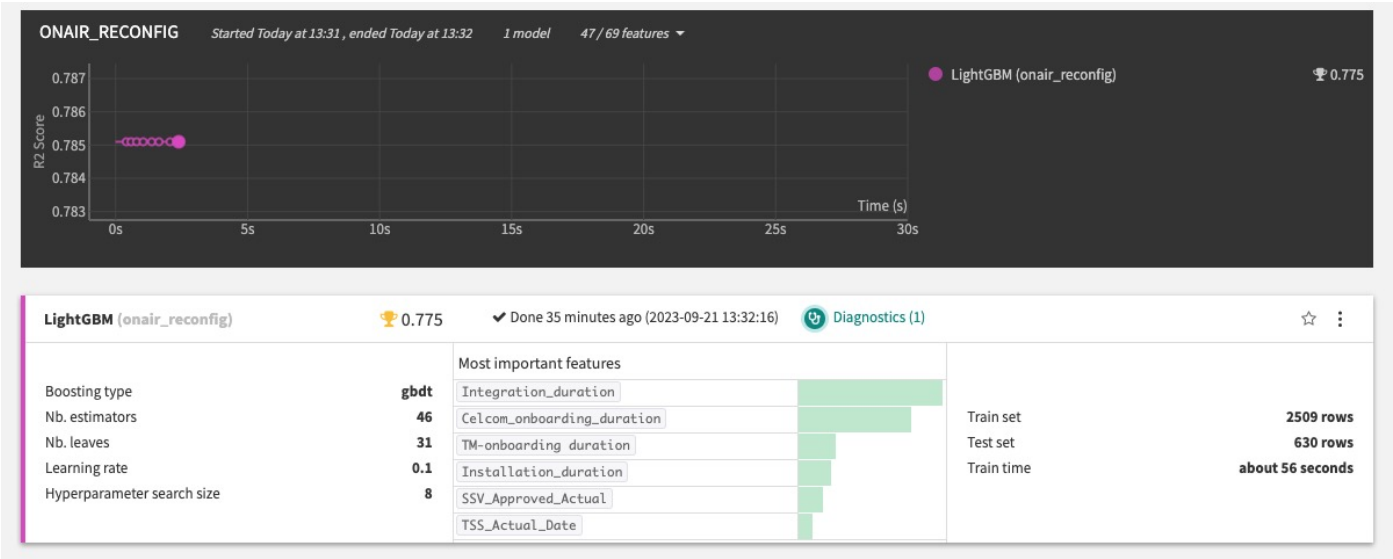
Deployment

- Once model is good enough decide what to do next:
- deploy the model
- collect more data and improve it
- revise business question again

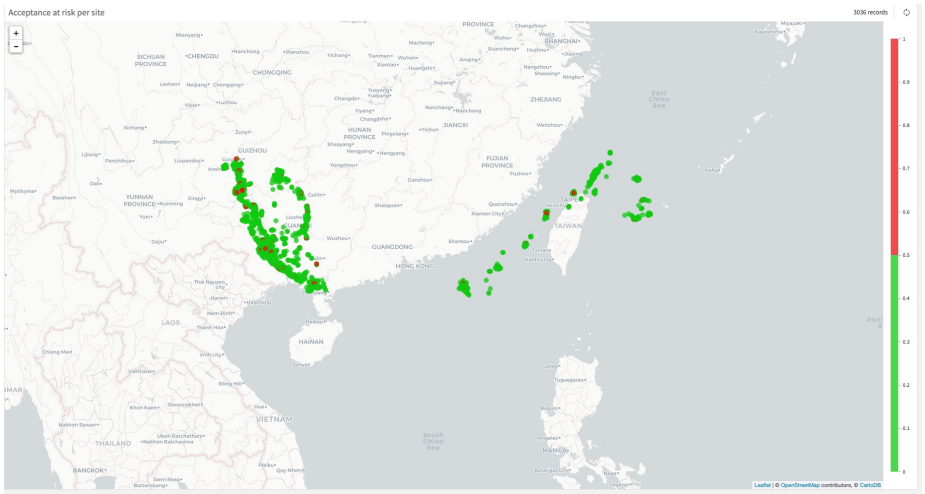


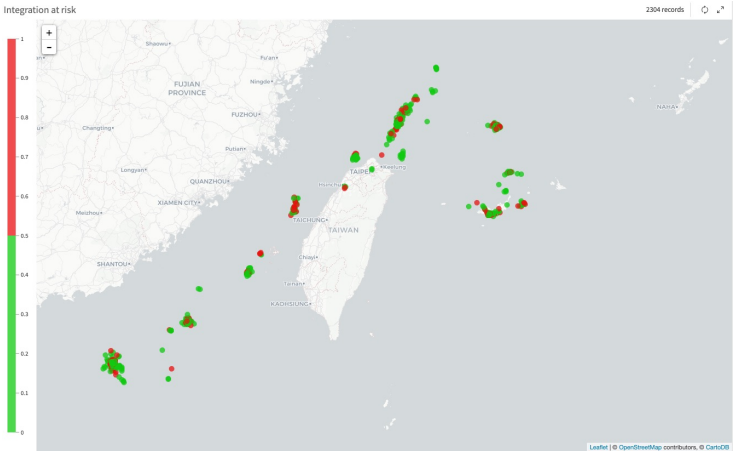
MLOps

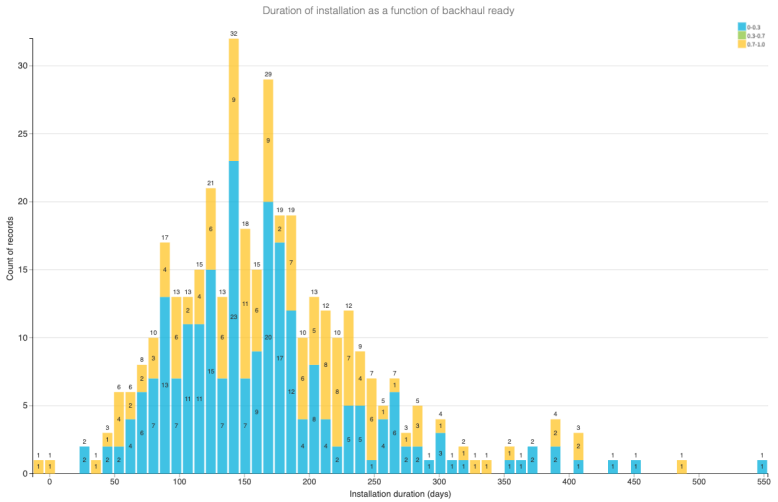
- Monitor the performance of the model because new data will be added to the dataset
- Retrain with new data



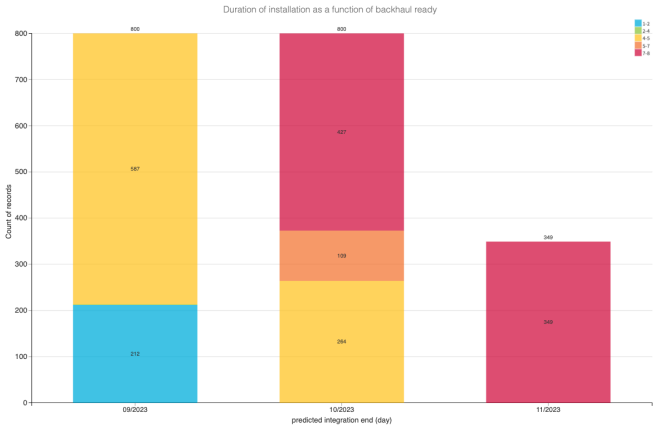
Business insights







Business insights




```
A function for prioritising the sites/rows.
'''

if row["Colocate"] == "Yes" and (row["CME_Plan_End_Date"] <= row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 1:
    #If it is co-located, CME is expected to be done and the backhaul is done: Priority 1
    return 1

elif row["Colocate"] == "No" and (row["CME_Plan_End_Date"] <= row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 1:
    #If it is not co-located, CME is expected to be done and the backhaul is done: Priority 2
    return 2

elif row["Colocate"] == "Yes" and (row["CME_Plan_End_Date"] > row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 1:
    #If it is co-located, CME is not expected to be done and the backhaul is done: Priority 3
    return 3

elif row["Colocate"] == "Yes" and (row["CME_Plan_End_Date"] <= row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 0:
    #If it is co-located, CME is expected to be done and the backhaul is required: Priority 4
    return 4

elif row["Colocate"] == "No" and (row["CME_Plan_End_Date"] > row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 1:
    #If it is not co-located, CME is not expected to be done and the backhaul is done: Priority 5
    return 5

elif row["Colocate"] == "Yes" and (row["CME_Plan_End_Date"] > row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 0:
    #If it is co-located, CME is not expected to be done and the backhaul is required: Priority 6
    return 6

elif row["Colocate"] == "No" and (row["CME_Plan_End_Date"] <= row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 0:
    #If it is not co-located, CME is expected to be done and the backhaul is required: Priority 7
    return 7

elif row["Colocate"] == "No" and (row["CME_Plan_End_Date"] > row['Predicted_integration_end_' + str(monthly_limit)]) and row["Backhaul_ready"] == 0:
    #If it is not co-located, CME is not expected to be done and the backhaul is required: Priority 8
    return 8

else:
    #Something was input incorrectly.
    #raise Exception("A value was input incorrectly")
    if row["Colocate"] == "Yes" and row["Backhaul_ready"] == 1:
        #If it is co-located, CME is not expected to be done and the backhaul is done: Priority 3
        return 3

    elif row["Colocate"] == "No" and row["Backhaul_ready"] == 1:
        #If it is not co-located, CME is not expected to be done and the backhaul is done: Priority 5
        return 5

    elif row["Colocate"] == "Yes" and row["Backhaul_ready"] == 0:
        #If it is co-located, CME is not expected to be done and the backhaul is required: Priority 6
        return 6

    elif row["Colocate"] == "No" and row["Backhaul_ready"] == 0:
        #If it is not co-located, CME is not expected to be done and the backhaul is required: Priority 8
        return 8
```

Summary

- 5G does not stop at the air interface
- besides RAN or core network technology other teams with other skillsets are necessary to enable a 5G communication link
- The rollout team is one of them (besides sourcing, product development, sales, finance...)
- Machine learning can help , f. ex. in prediting the sites that are at risk of not being finished on time, thus incurring a penalty
- Machine learning can help reproritise the workforce for better use of them
- Domain knowledge and dashboarding are two of the most important assets in order to deliver a successful ML use case