Digitalized Logistics

Advanced Analytics@POST AG

Christoph Bodner & Thomas Laber

AGENDA

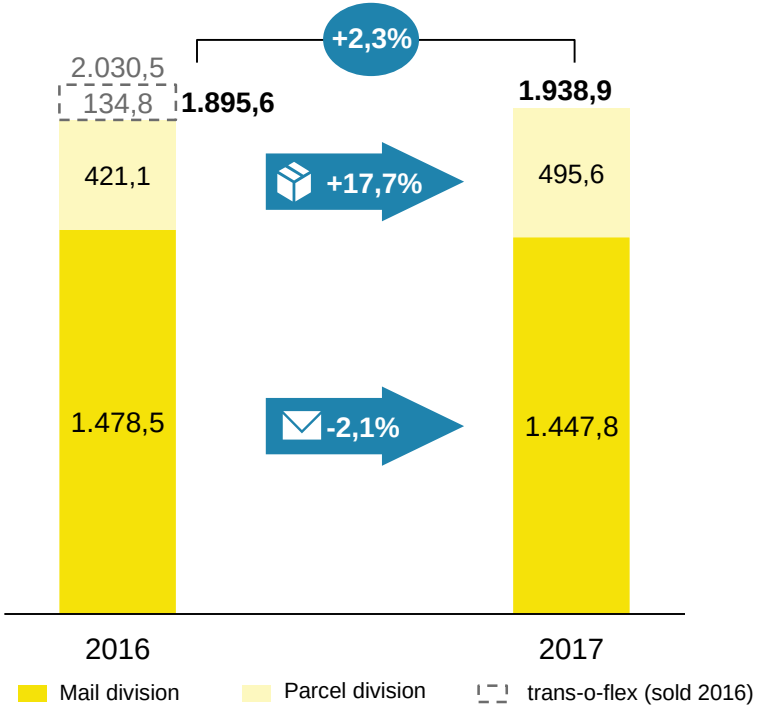
	01	02	03
Topics	Who we are (obligatory marketing stuff…)	What we do	How we do it
	Data Science@Post AG: - Overview: Post AG - Interdisciplinary team	Projects we work on: - Parcel volume forecast - Expected delivery times	With a combination of: - Math/statistics - Coding (Azure + R) - Grit & perseverance☺

AGENDA

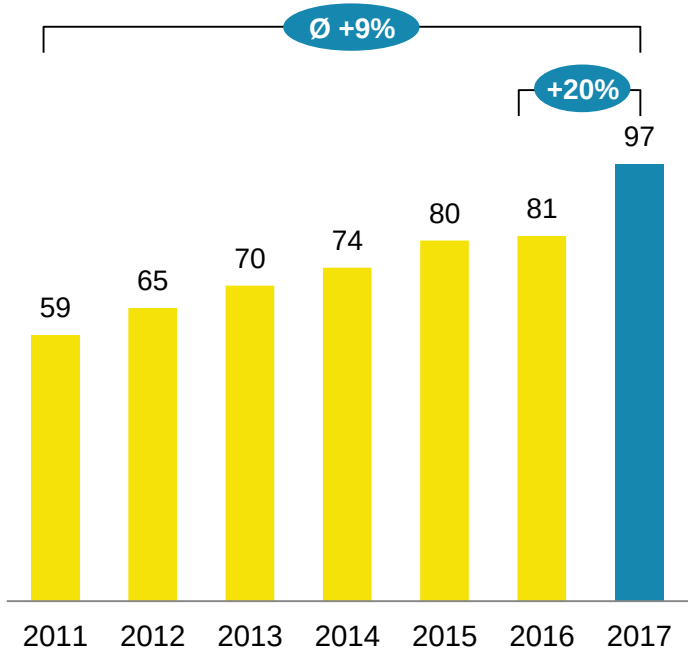
	01	02	03
Topics	Who we are (obligatory marketing stuff…)	What we do	How we do it
	Data Science@Post AG: - Overview: Post AG - Interdisciplinary team	Projects we work on: - Parcel volume forecast - Expected delivery times	With a combination of: - Math/statistics - Coding (Azure + R) - Grit & perseverance☺

OVERVIEW: POST AG
COMPANY PERFORMANCE & PARCEL VOLUMES OVER TIME

SALES PERFORMANCE
mio EUR



PARCEL VOLUMES OF AUSTRIAN POST
mio parcels



OVERVIEW: POST AG
STRONG PRESENCE IN EASTERN EUROPE

GROWTH FOCUS ON PACKAGE & LOGISTICS DIVISION

OUR TEAM
PEOPLE WHO LIKE $\pi z^2 a$ IN EVERY FORM☺

Christoph Bodner

Lead Data Scientist

Quantitative Finance (WU)
Prev.: KPMG

Thomas Laber

Senior Data Scientist

Business Informatics (TU)
Prev.: Accenture

Martin Blöschl

Junior Data Scientist

Computational Intelligence (TU)

Raphael Pesl

Junior Data Scientist

Mathematics (TU)

AGENDA

	01	02	03
Topics	Who we are (obligatory marketing stuff…)	What we do	How we do it
	Data Science@Post AG: - Overview: Post AG - Interdisciplinary team	Projects we work on: - Parcel volume forecast - Expected delivery times	With a combination of: - Math/statistics - Coding (Azure + R) - Grit & perseverance☺

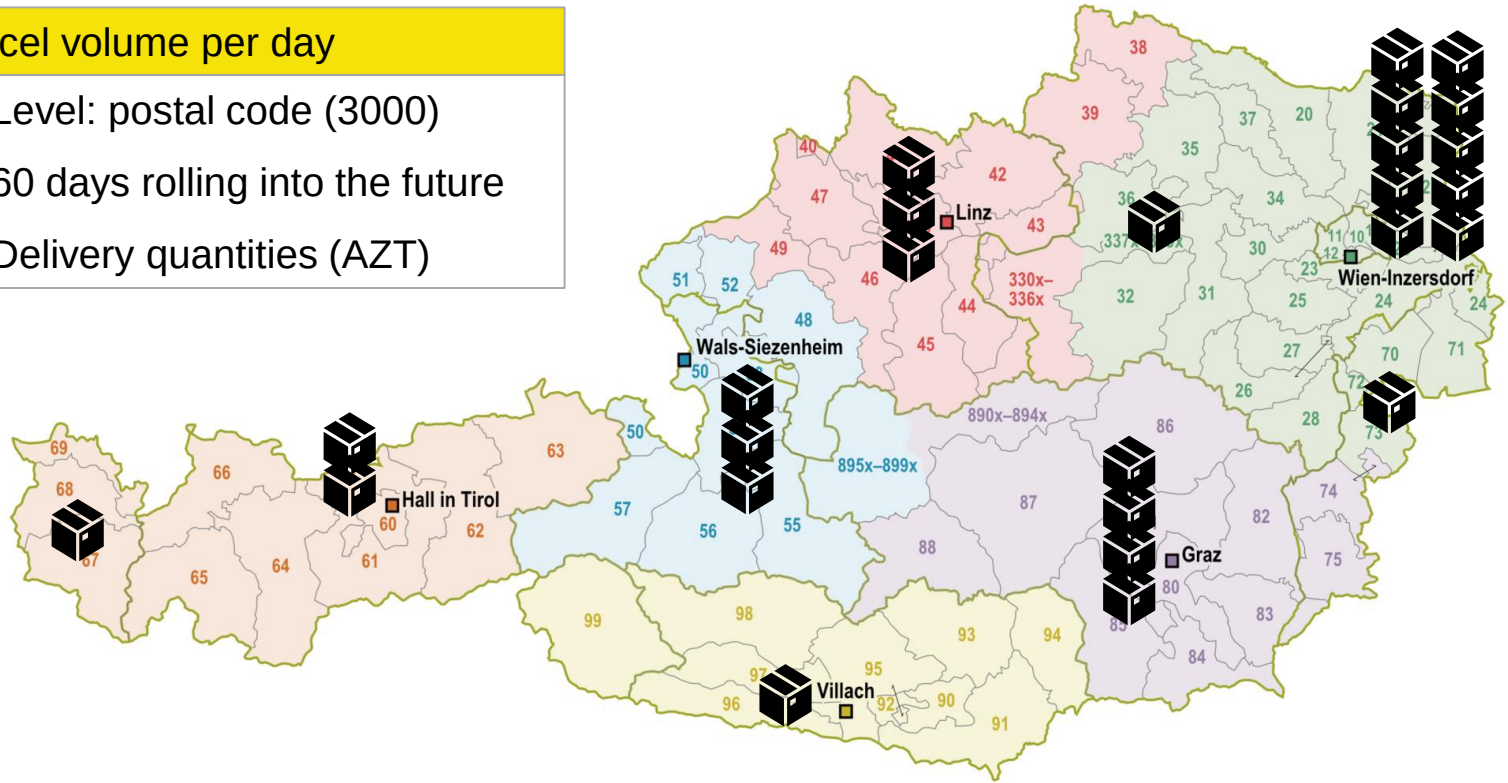
Parcel Volume: Daily Forecast

PARCEL VOLUME FORECAST

HOW MANY PARCELS WILL WE NEED TO DELIVER IN THE FUTURE?

Parcel volume per day

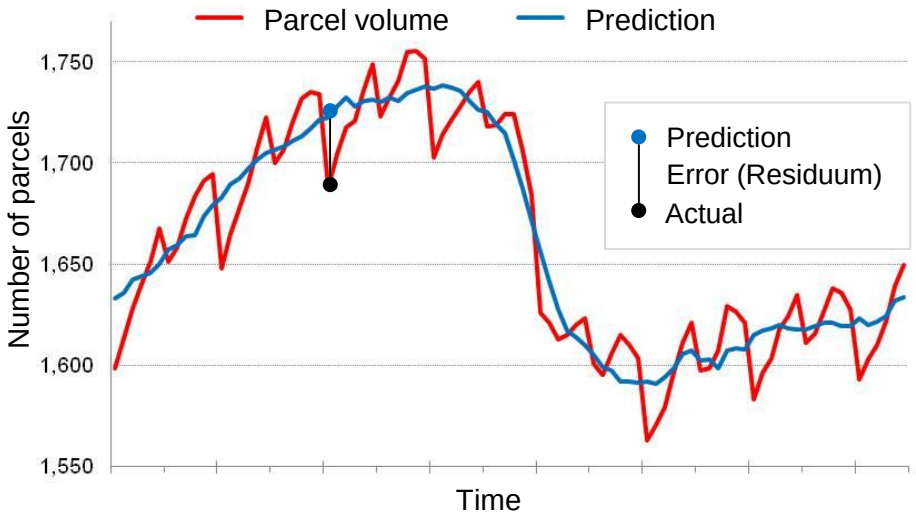
- Level: postal code (3000)
- 60 days rolling into the future
- Delivery quantities (AZT)



THE PROBLEM IS SIMPLE TO FORMULATE …
… BUT NOT THAT EASY TO SOLVE

Post

Parcel volume over time



As small as possible

Package volume = prediction + error

h(X)

How can we find the optimal h(X)?

$$\hat{h} = arg \min_{h \in H} R_{emp}(h)$$

where:

$$R_{emp}(h) = \frac{1}{m} \sum_{i=1}^{m} L(h(x_i), y_i)$$

SO THE QUESTION IS:
HOW DO WE FIND THE OPTIMAL PREDICTION FUNCTION?

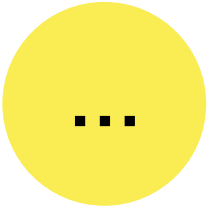
There are lots of different ways to find h(X):

Neural networks

Linear Regression
(Image is no mistake ☺)

Gradient Boosting
(with decision trees)

currently in use



other models

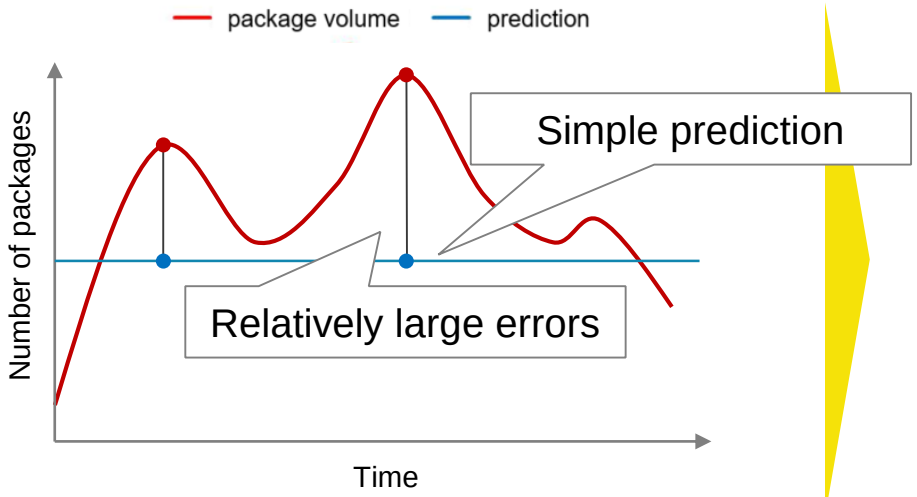
BUT NEURAL NETS ARE SO SEXY!
WHY USE GRADIENT BOOSTING?

Gradient Boosting offers a very good combination of complexity & performance

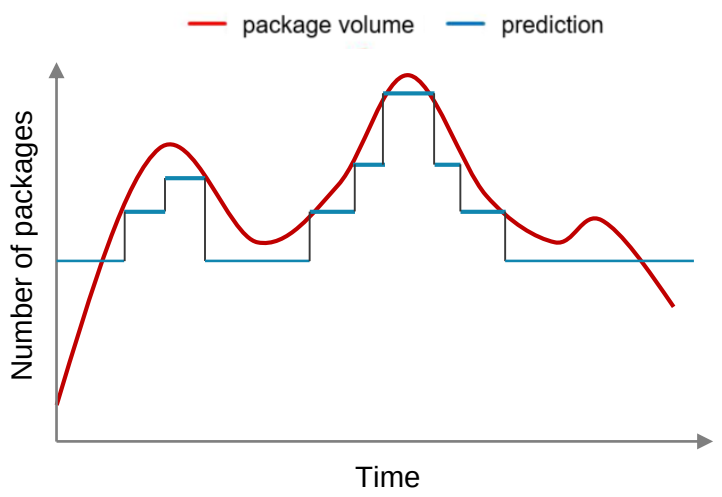
Source: Toyota + Black Hawk: blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/

LET‘S TAKE A DEEPER LOOK
HOW DOES GRADIENT BOOSTING WORK?

Start with simple prediction ...

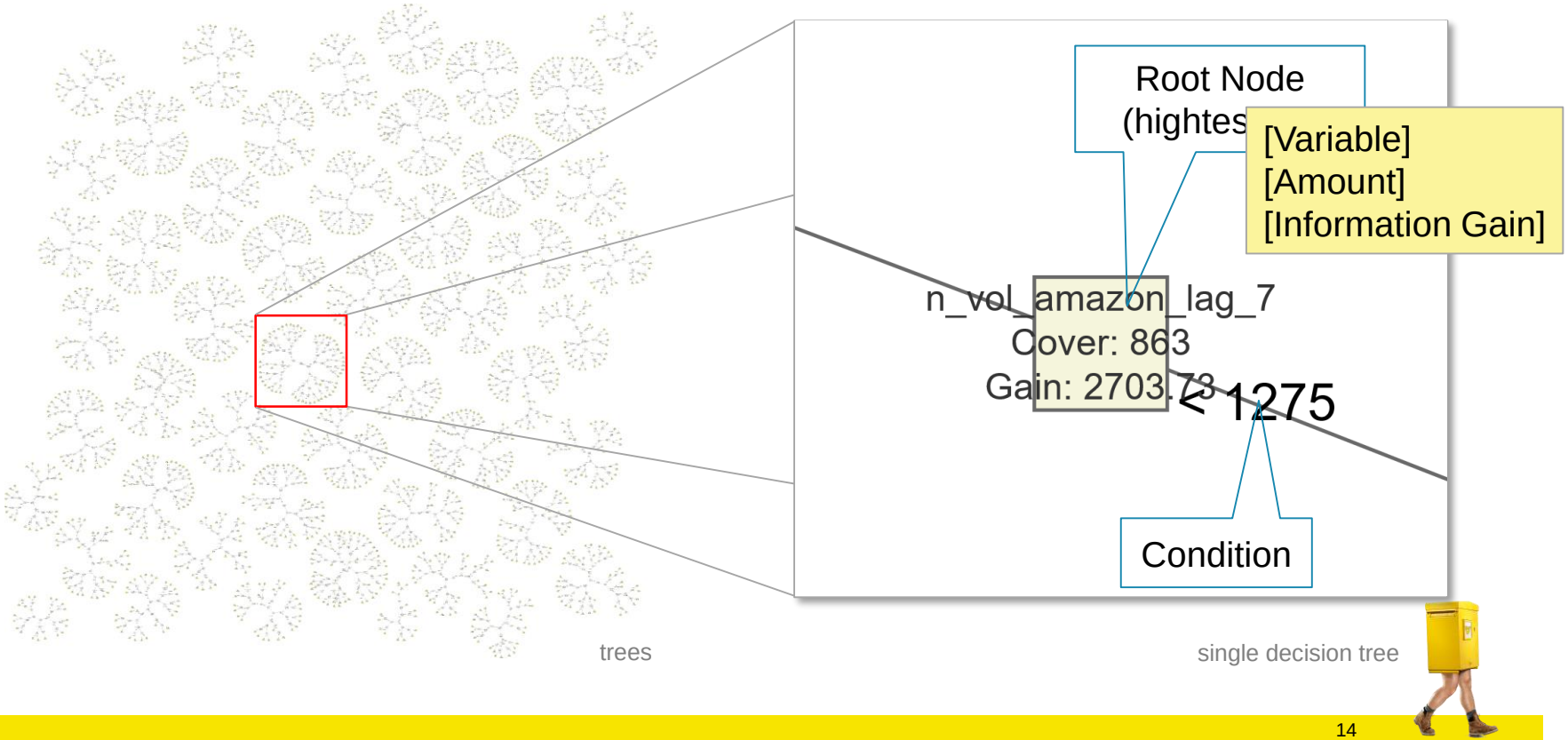


... and gradually improve

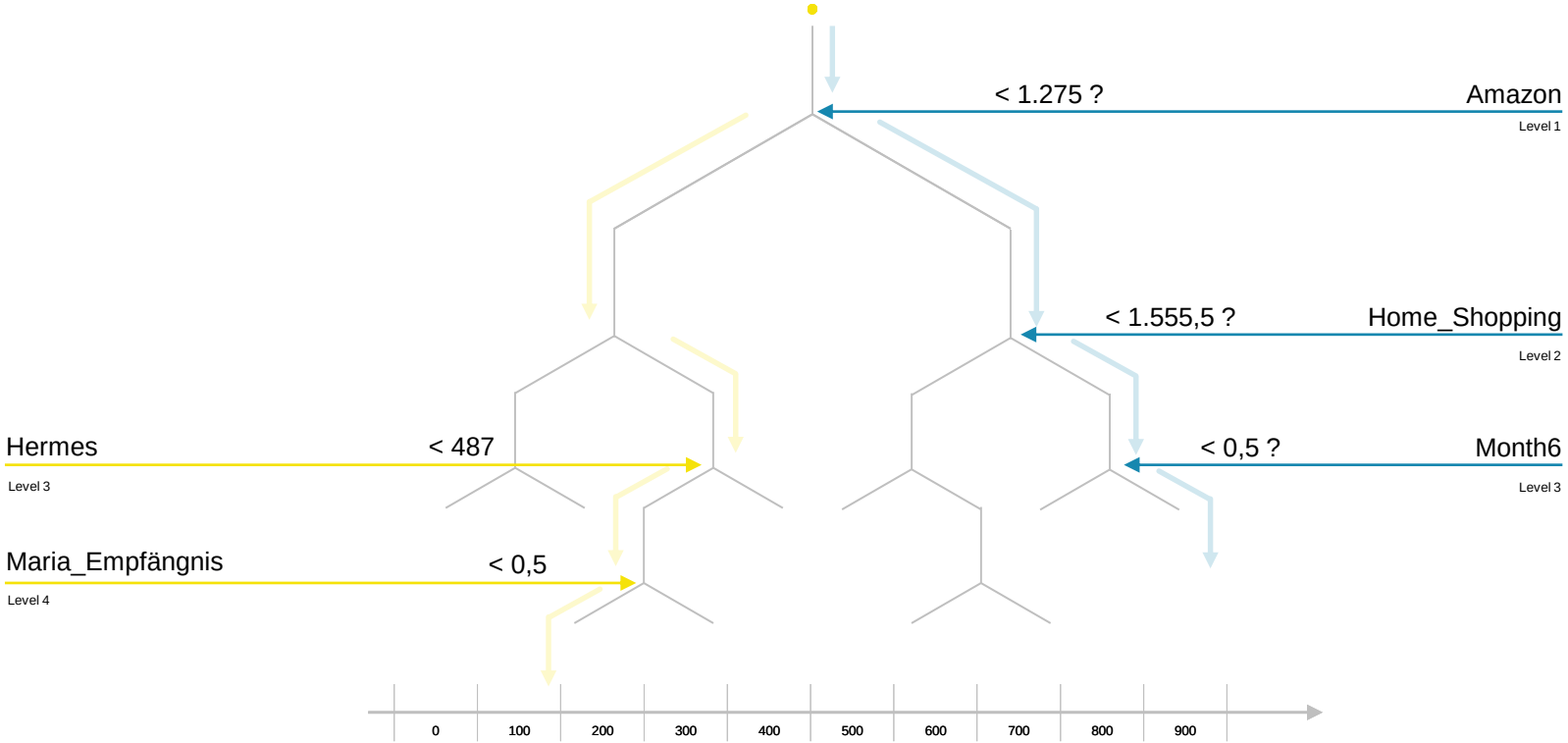


Individual weak forecasts are combined to form a strong prediction

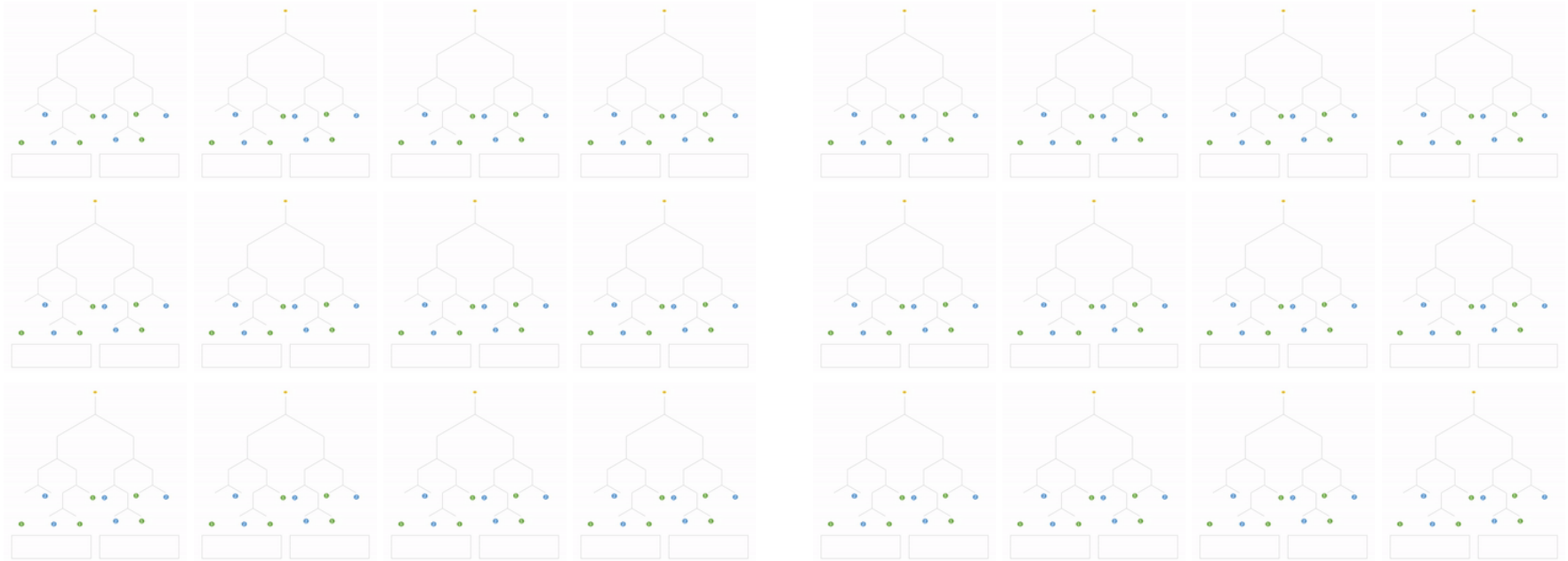
XGBOOST - OVERVIEW
VISUALISATION



XGBOOST
A CLOSER LOOK



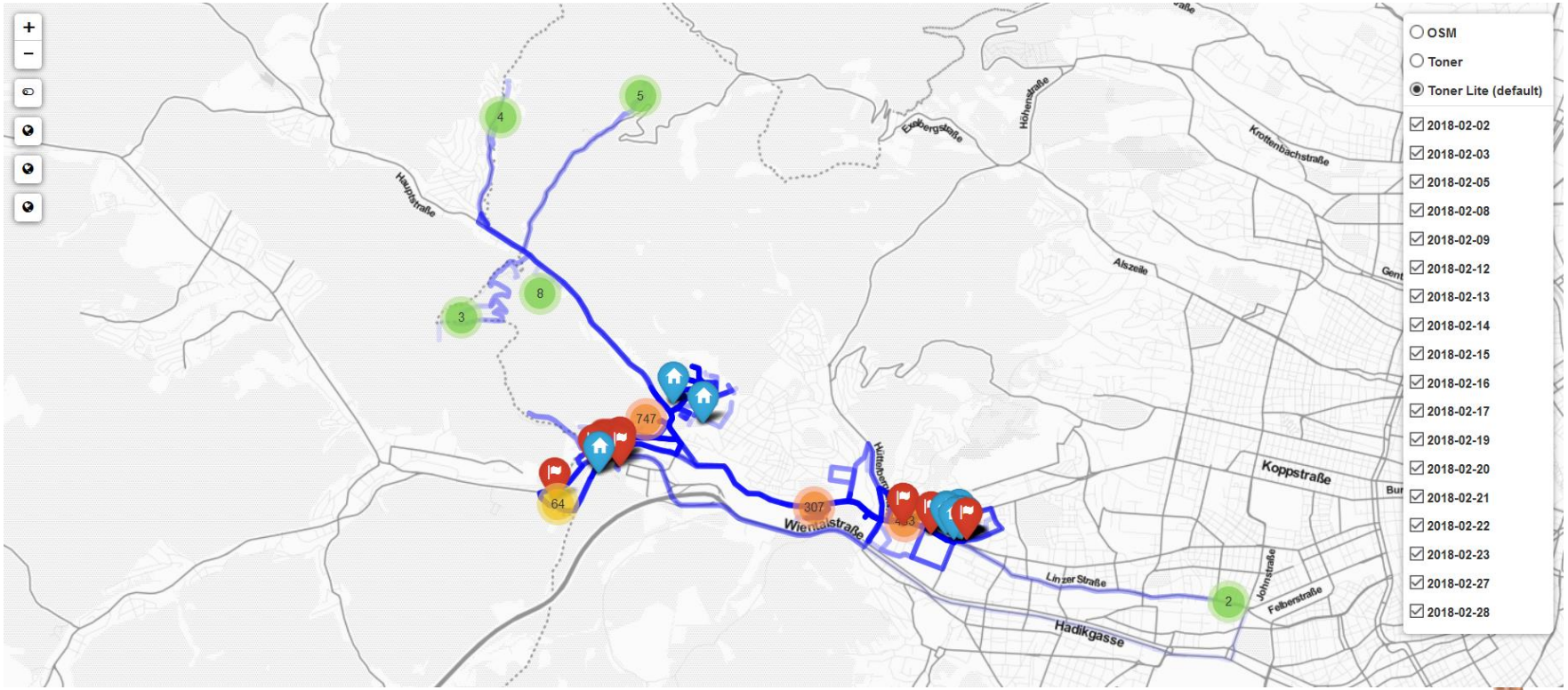
XGBOOST
A CLOSER LOOK



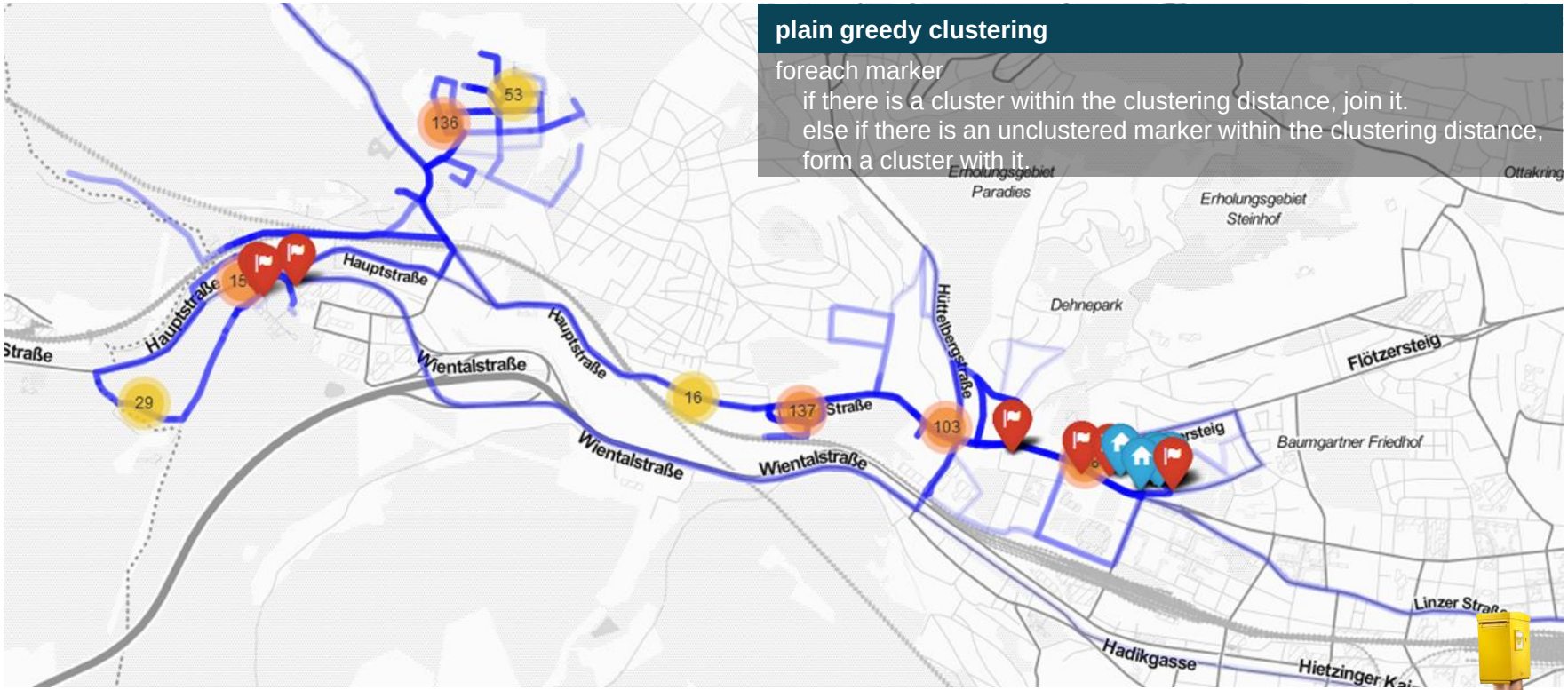
Parcel Delivery: 1h-Time Window

PREDICTING PARCEL DELIVERY TIMES

PROBLEM OVERVIEW



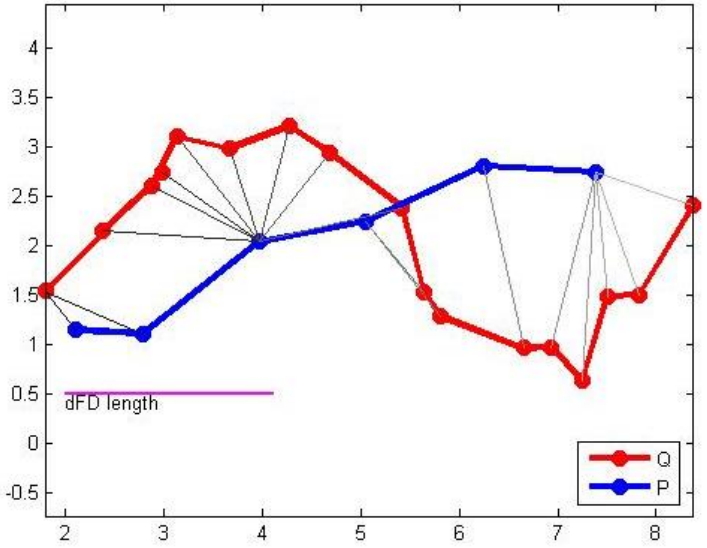
MOSTLY VERY CONSTANT TOURS
DELIVERY ‘CLUSTERS‘ CLEARLY VISIBLE

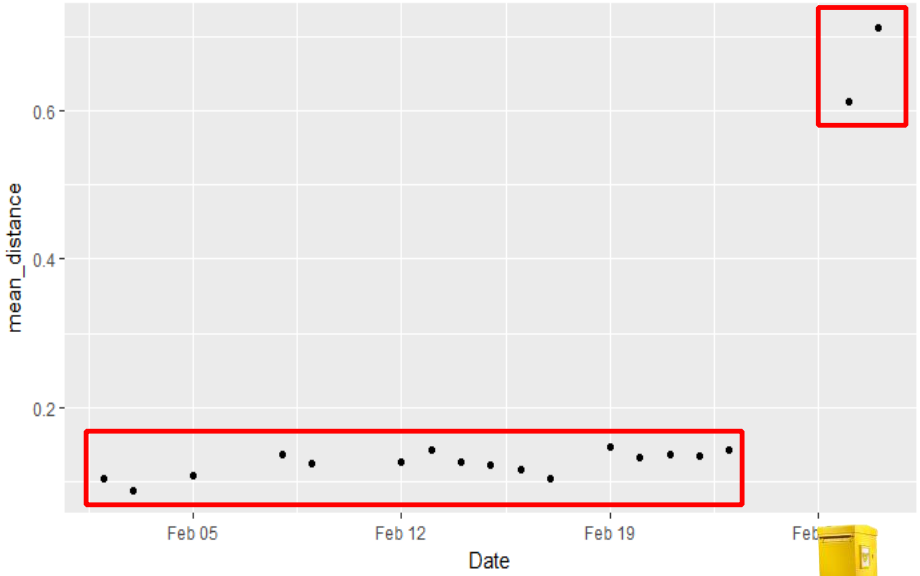


SOMETIMES DRASTIC DIFFERENCES
TOUR CAN CHANGE RAPIDLY

Using Frechet-distance to compare tours …

... shows rapid change from one day to next

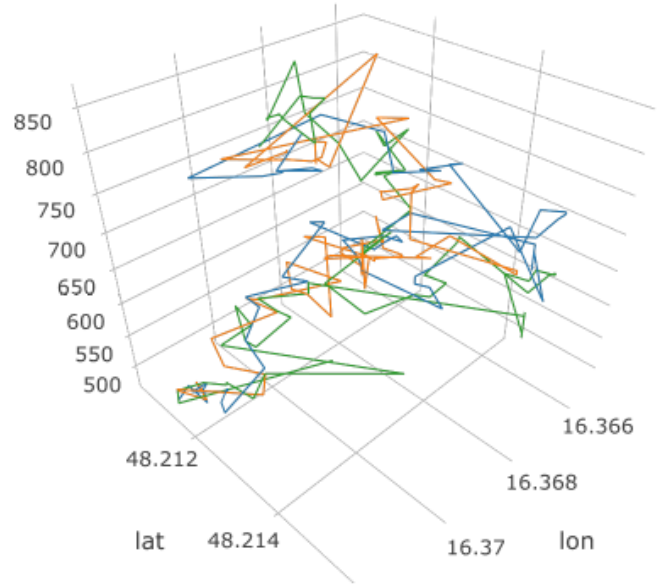




Source: http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.pdf
https://www.rdocumentation.org/packages/kmlShape/versions/0.9.5/topics/distFrechet

HOW CAN WE PREDICT DELIVERY TIMES?
IDEA: FIND ‘SIMILAR’ ROUTES IN THE PAST

Routes in location-time matrix (part of Vienna)



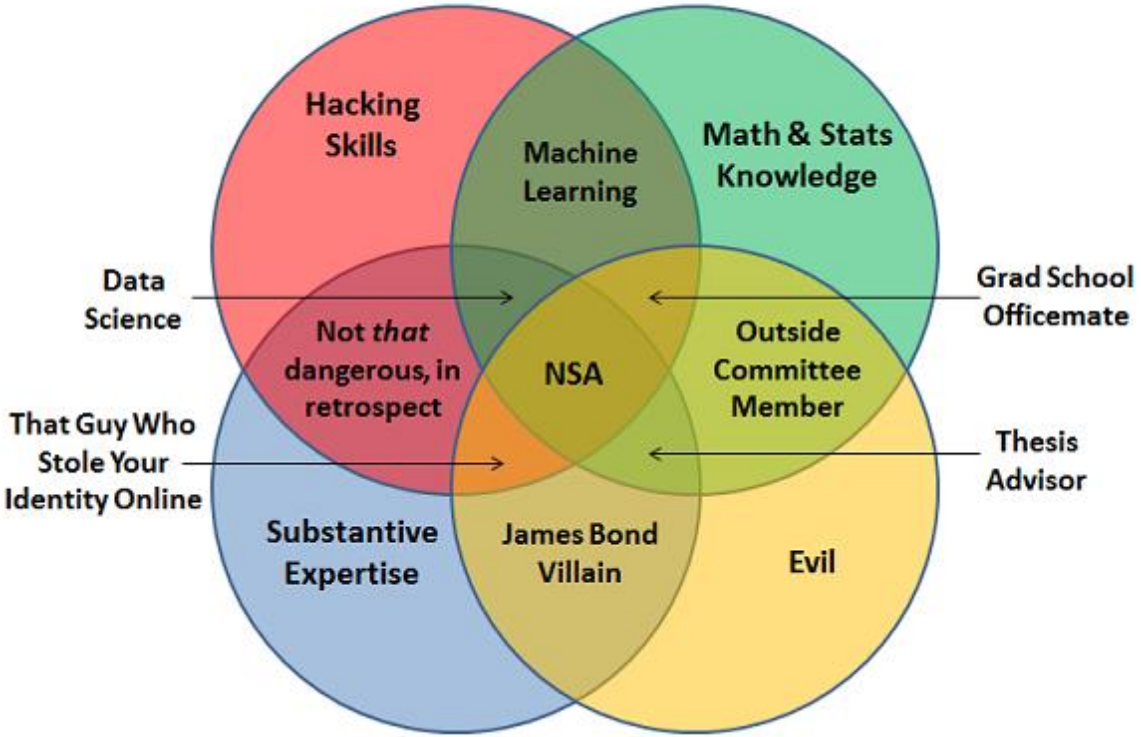
When are two routes similar?

- Similar trajectory, but differing stops and time points
- Similar trajectory, similar stops and differing time points
- Similar trajectory, similar stops and similar time points

AGENDA

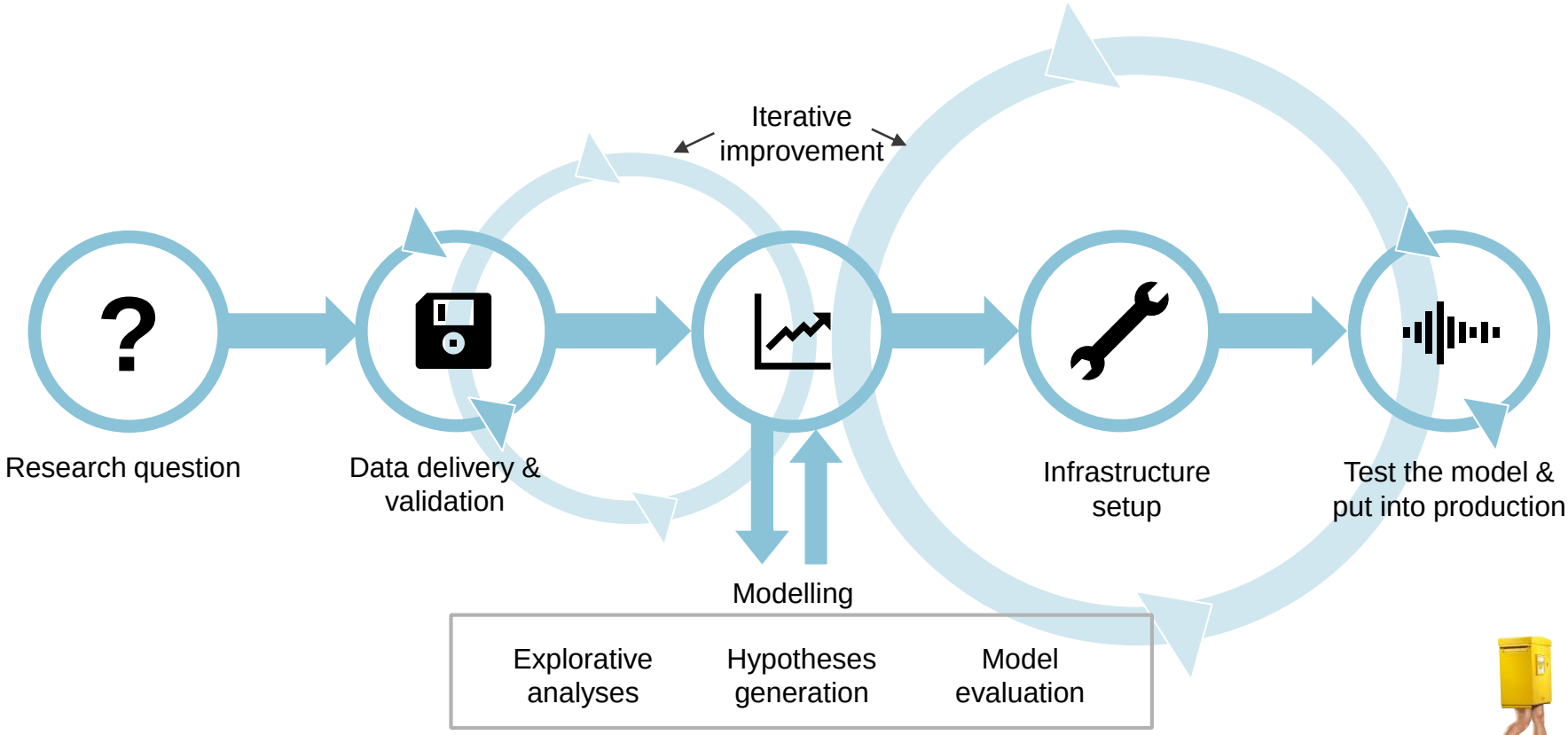
	01	02	03
Topics	Who we are (obligatory marketing stuff…)	What we do	How we do it
	Data Science@Post AG: - Overview: Post AG - Interdisciplinary team	Projects we work on: - Parcel volume forecast - Expected delivery times	With a combination of: - Math/statistics - Coding (Azure + R) - Grit & perseverance☺

DATA SCIENTISTS NEED MANY SKILLS
ASKING THE RIGHT QUESTIONS ≥ ALGORITHMS

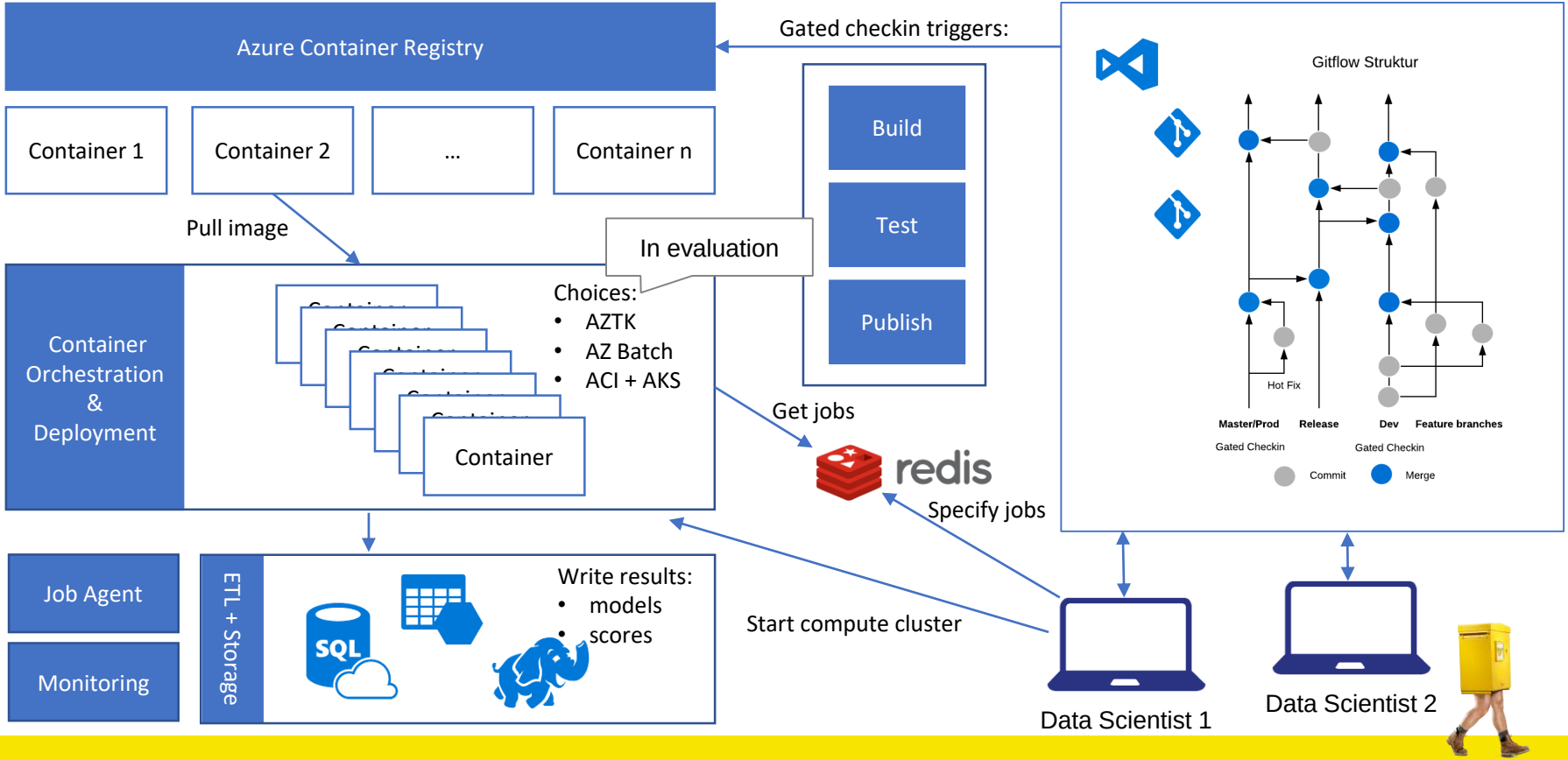


Source: http://joelgrus.com/2013/06/09/post-prism-data-science-venn-diagram/

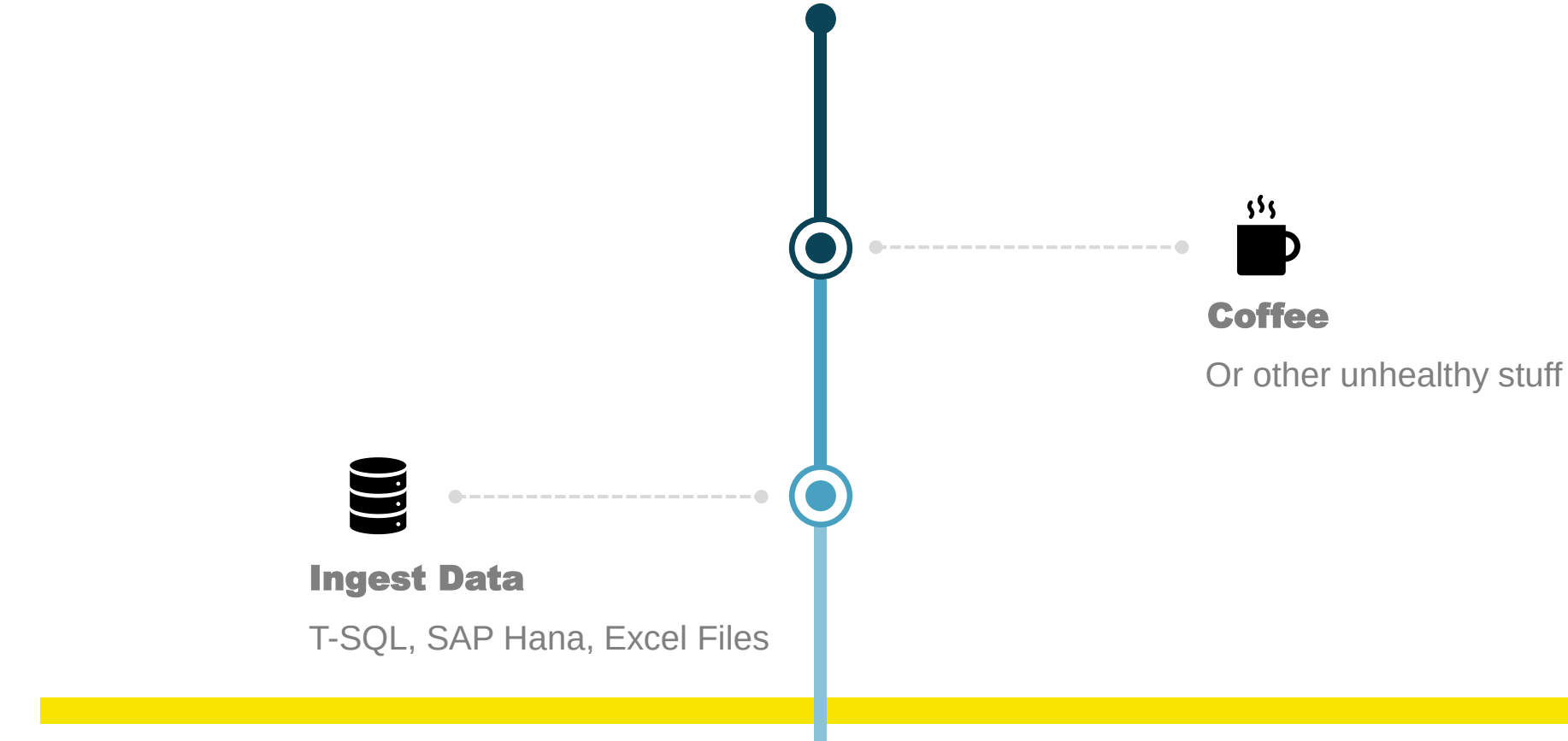
PROCESS OVERVIEW
DATA SCIENCE PROJECTS ARE ALWAYS “AGILE”



OUR STACK
WE ARE CURRENTLY BUILDING A HPC ENVIRONMENT



A Day in the Life of a Data Scientist

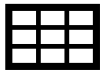

Coffee
Or other unhealthy stuff
Ingest Data
T-SQL, SAP Hana, Excel Files

Explore Data

Test hypotheses

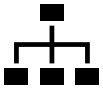
Data Validation

Done in R or Python



Feature Engineering

Identify relevant variables
or create them



Deployment

Does is scale?
Are we ready for the cloud?
Is the cloud ready for us?

Prototyp Model

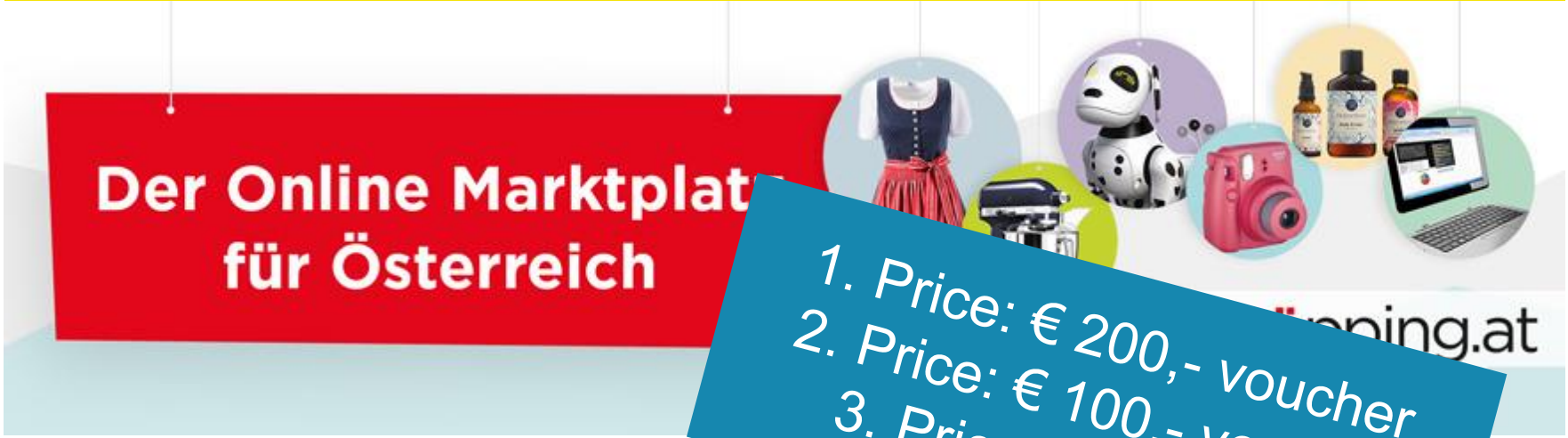
create a baseline
Scrum Demo

LET‘S TAKE A LOOK…

Data Science Challenge

PREDICTION CHALLENGE
FORECAST & WIN!

Predict 1h-delivery-time windows for our customer parcels



All the information you need:

https://github.com/POSTAG/time_window_prediction

THANK YOU FOR YOUR ATTENTION!