SINGULARITY
TECH DAY

#STechDay2020

Lykeion, one ML platform to rule them all

SINGULARITY
TECH DAY

#STechDay2020

ORGANIZATION

plain concepts

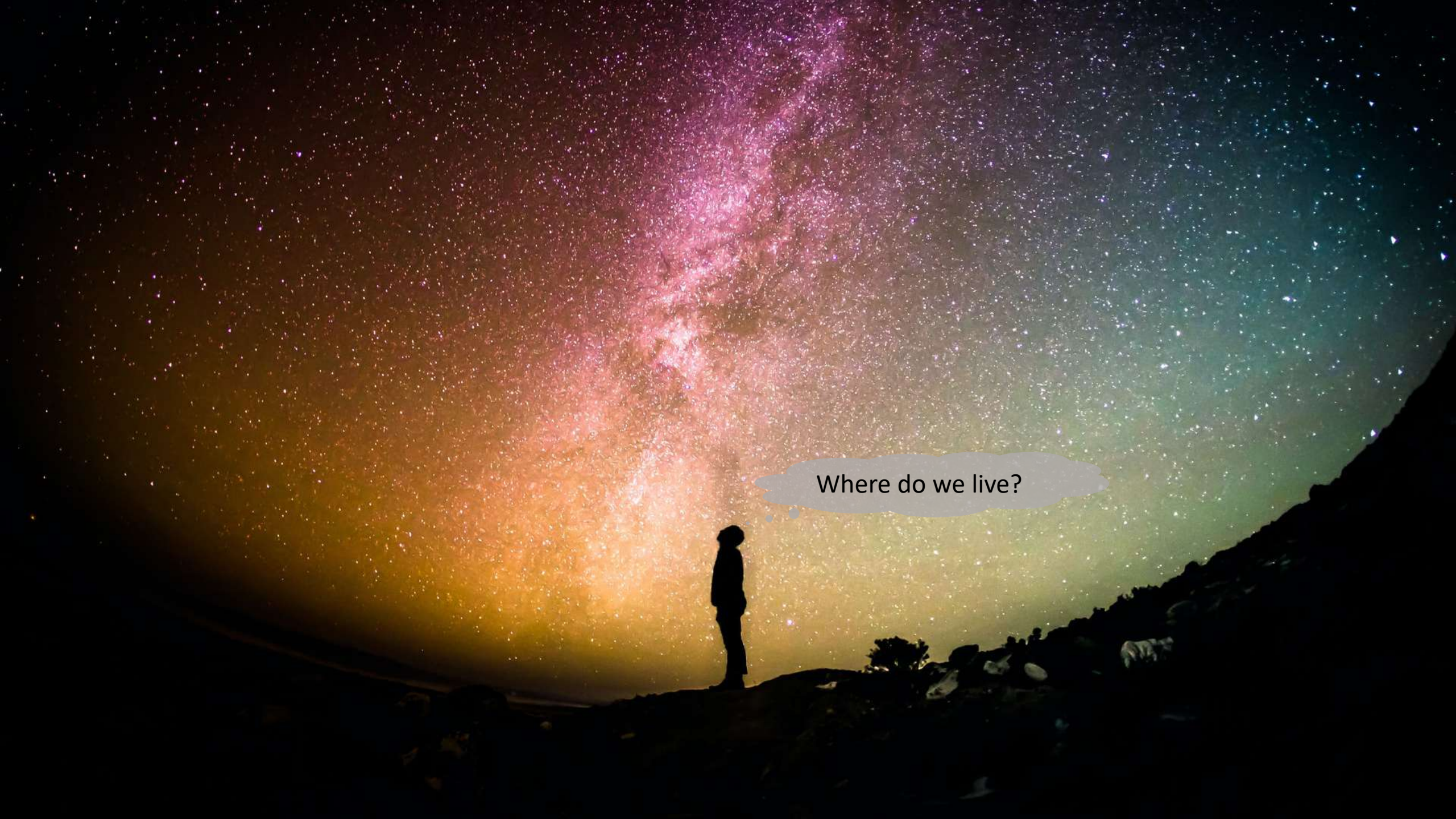
SPONSORS

Microsoft

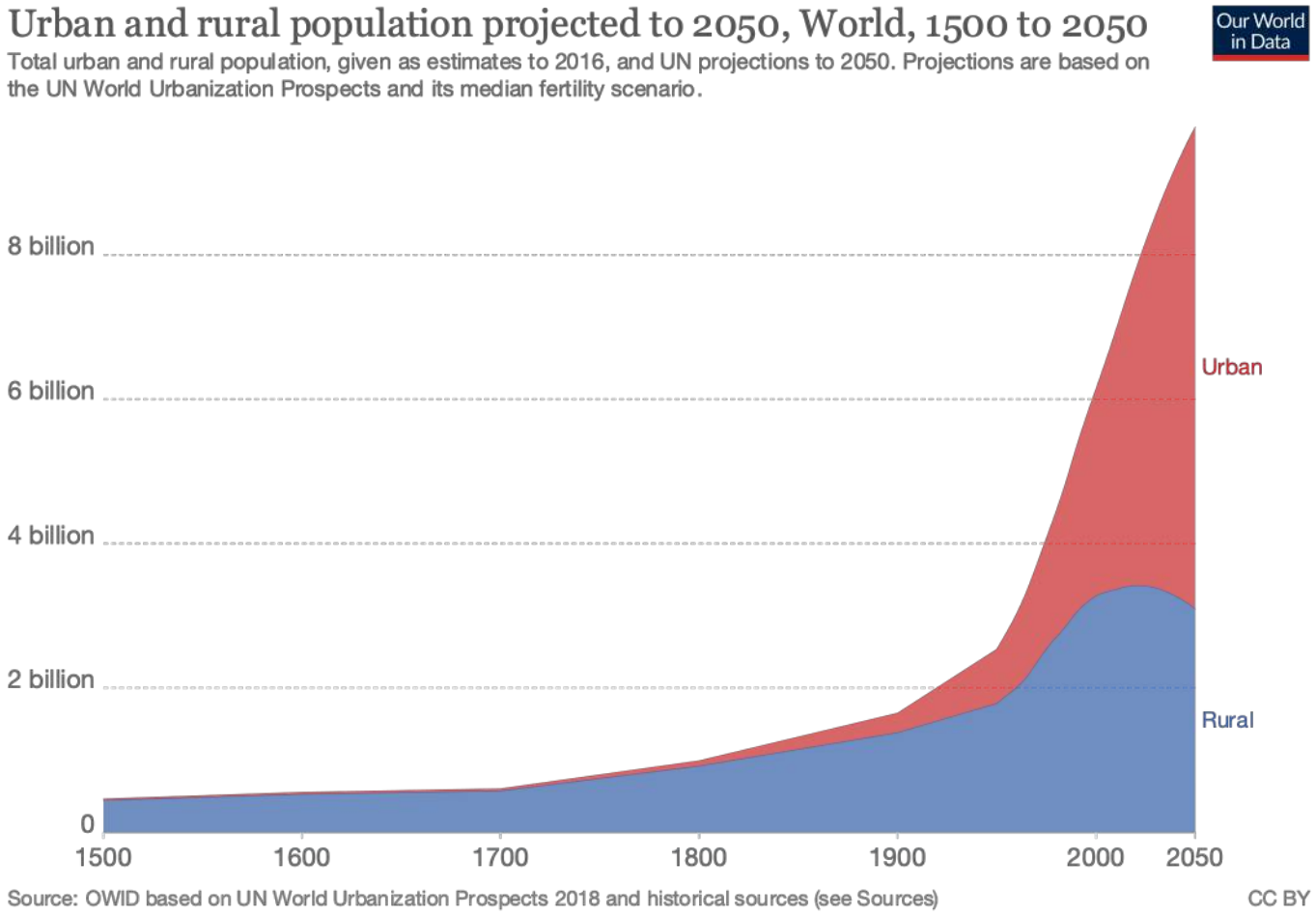
SUPPORT

World Central Kitchen

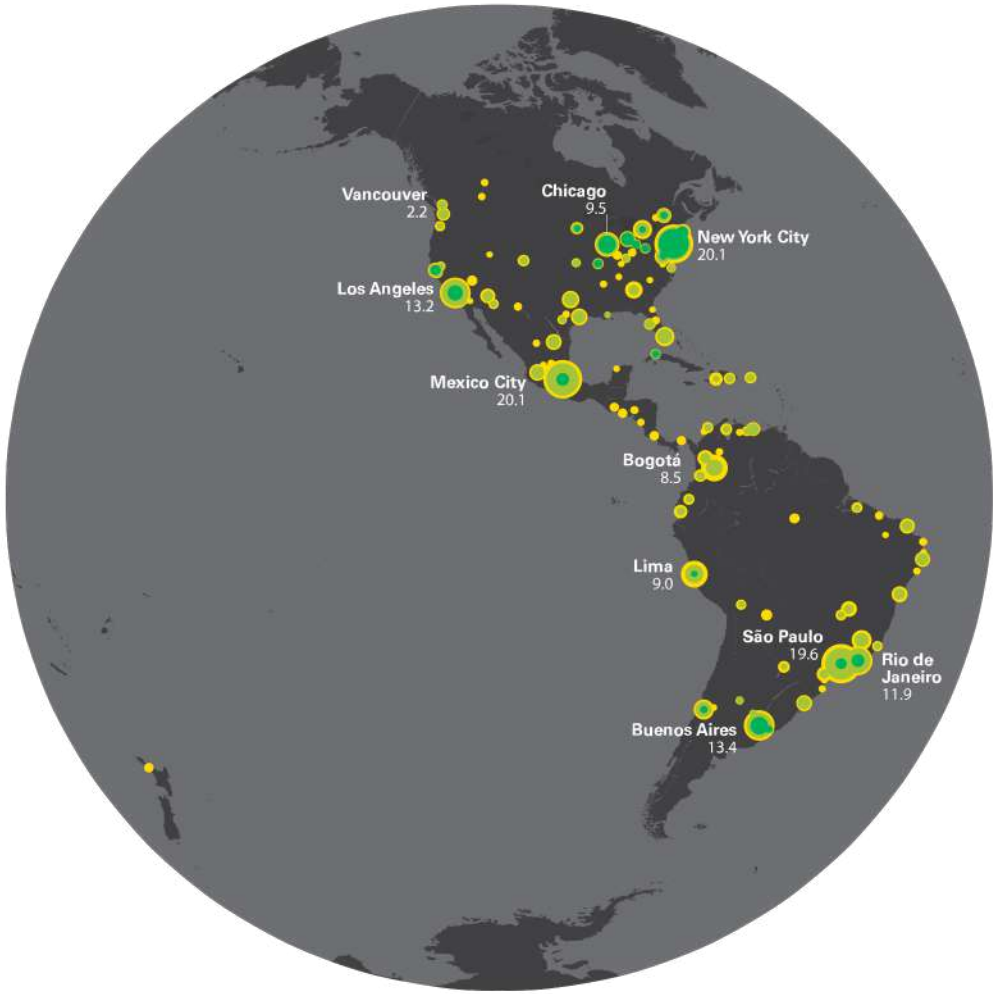
THANK YOU!

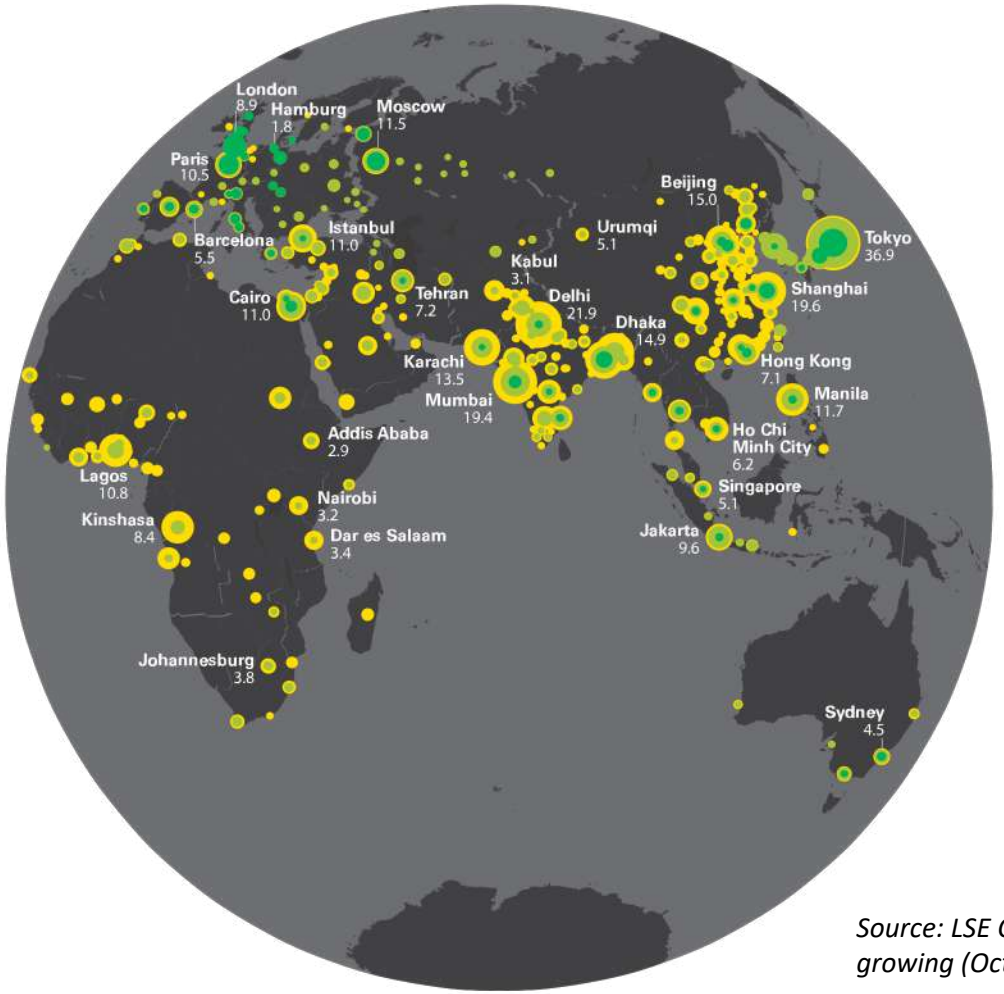

Where do we live?

Where do we live?



How are cities growing?





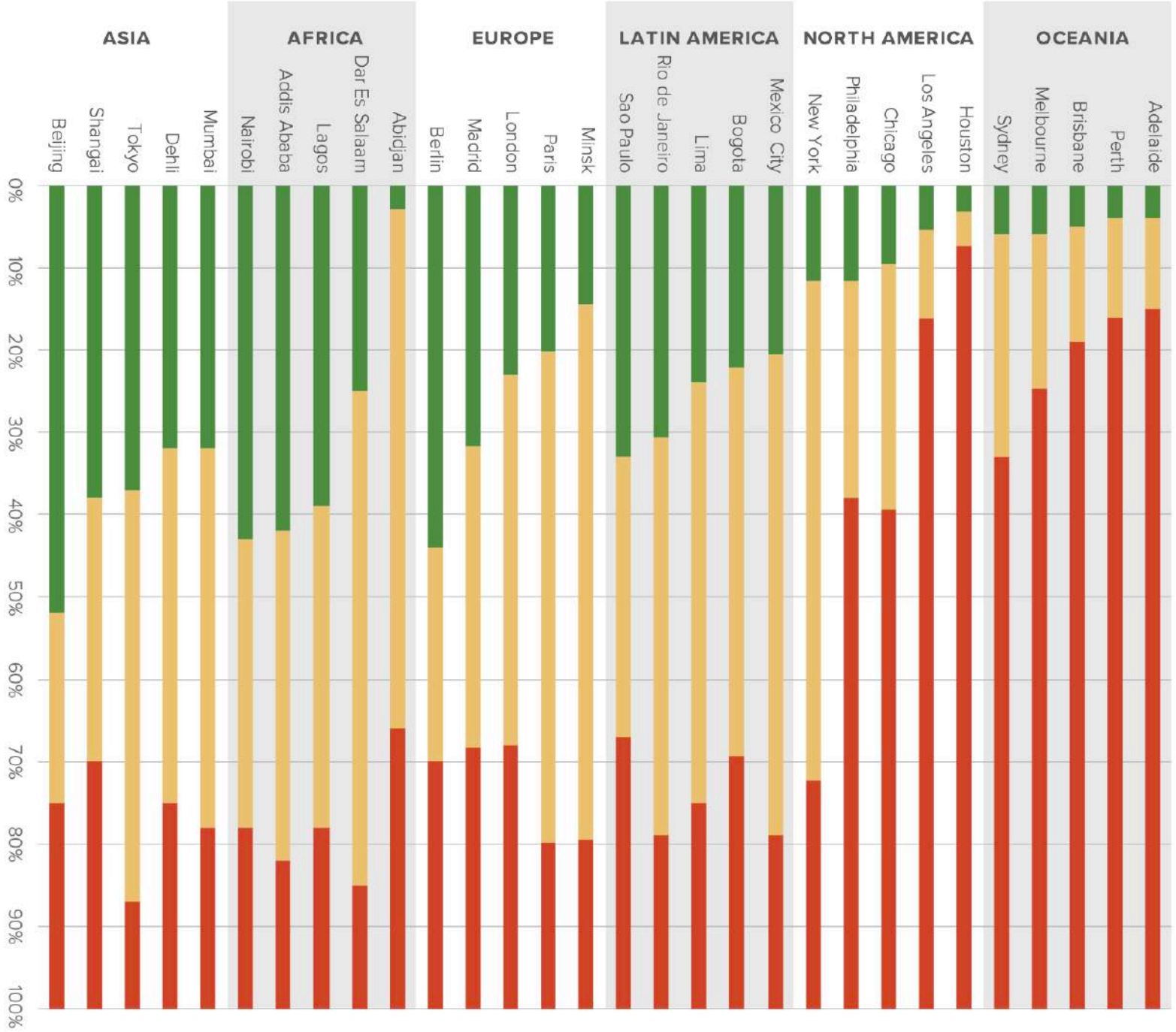
Metropolitan population in millions (2010)

- 1950
- 1990
- 2025

Source: LSE Cities. Where cities are growing (October 2013)

How all these people travel around the city?

Modal share in each region for the 5 largest cities



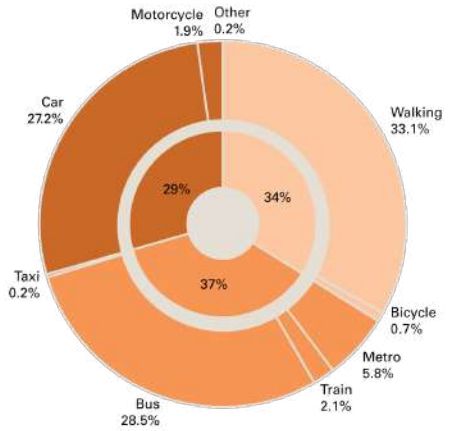
- Non-motorized Transport
- Public Transport
- Private Cars and Others

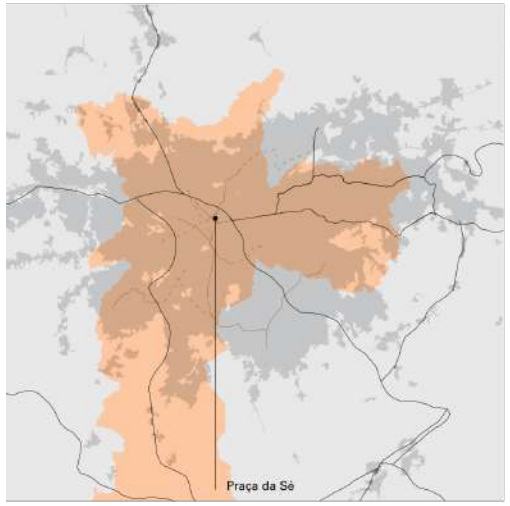
Source: Coalition for Urban Transitions

SINGULARITY
TECH DAY

cabify
#STechDay2020

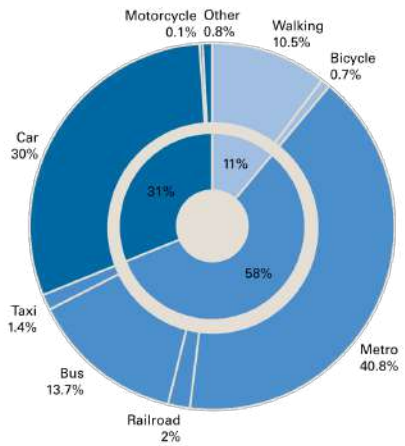
SÃO PAULO

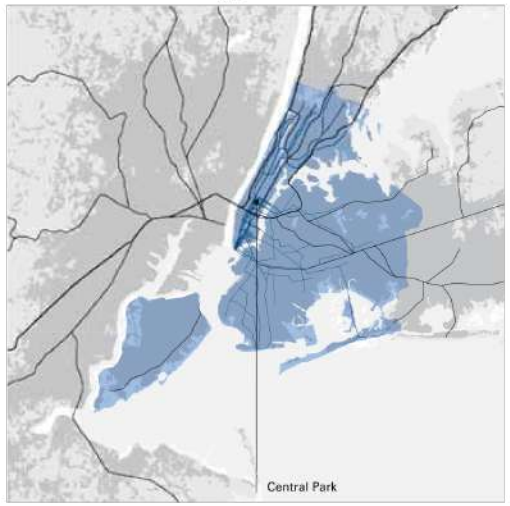




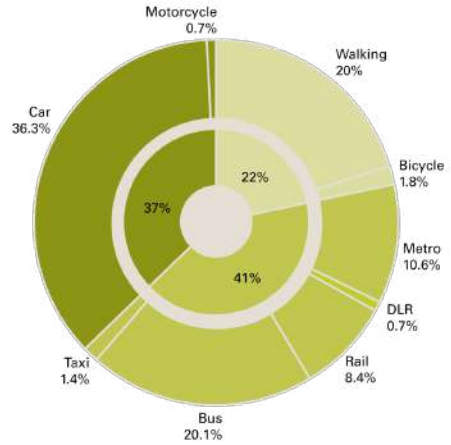
NEW YORK CITY

* The modal shares below refer to trips to work only.

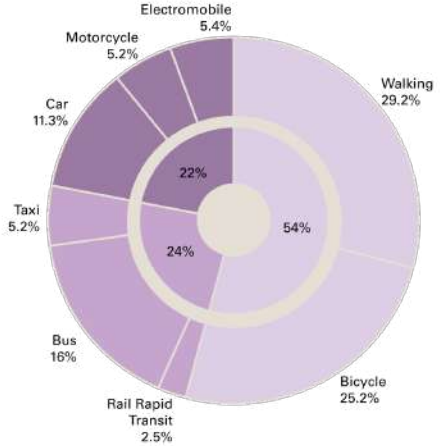


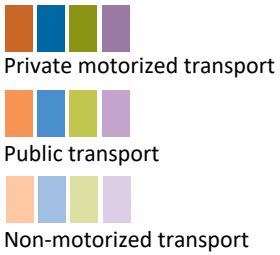


LONDON



SHANGHAI

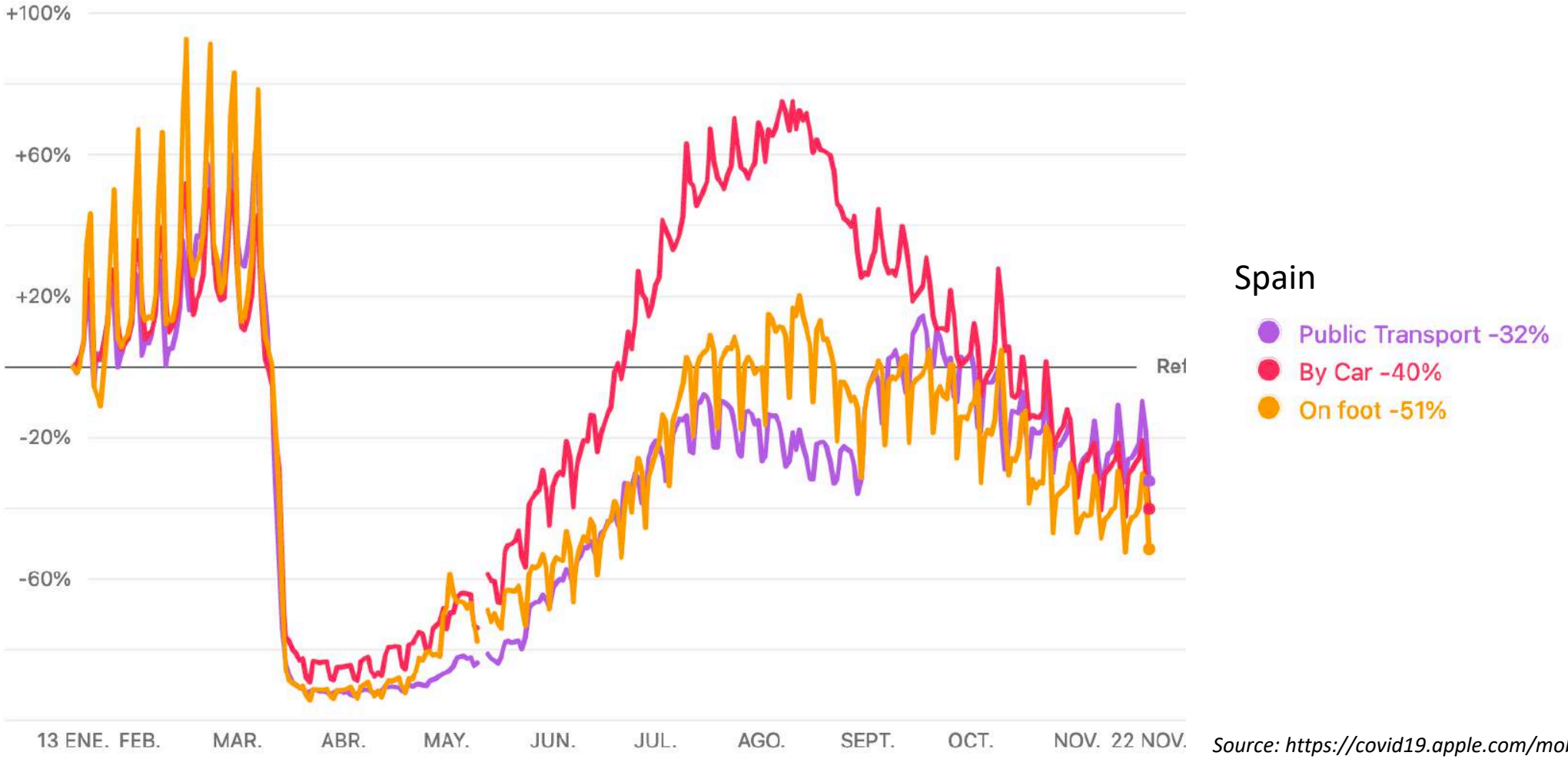




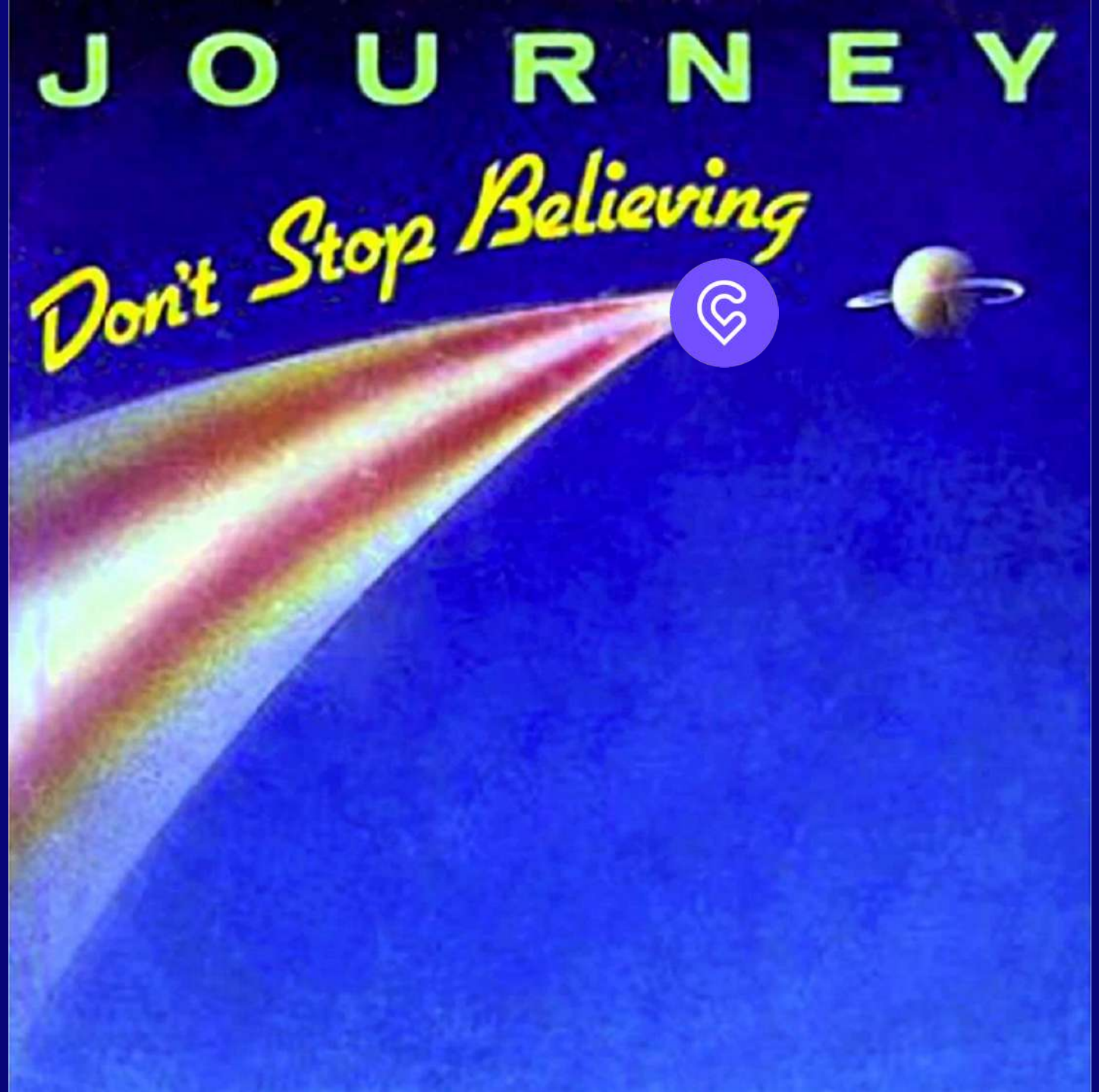
Source: LSE Cities. Where cities are growing (October 2013)

COVID-19 Arrived

Change in mobility paradigm



Source: https://covid19.apple.com/mobility


JOURNEY
Don't Stop Believing

SINGULARITY
TECH DAY

cabify
#STechDay2020

Eduardo Matallanas

Senior Data Scientist @ Cabify

Knowmad interested in:

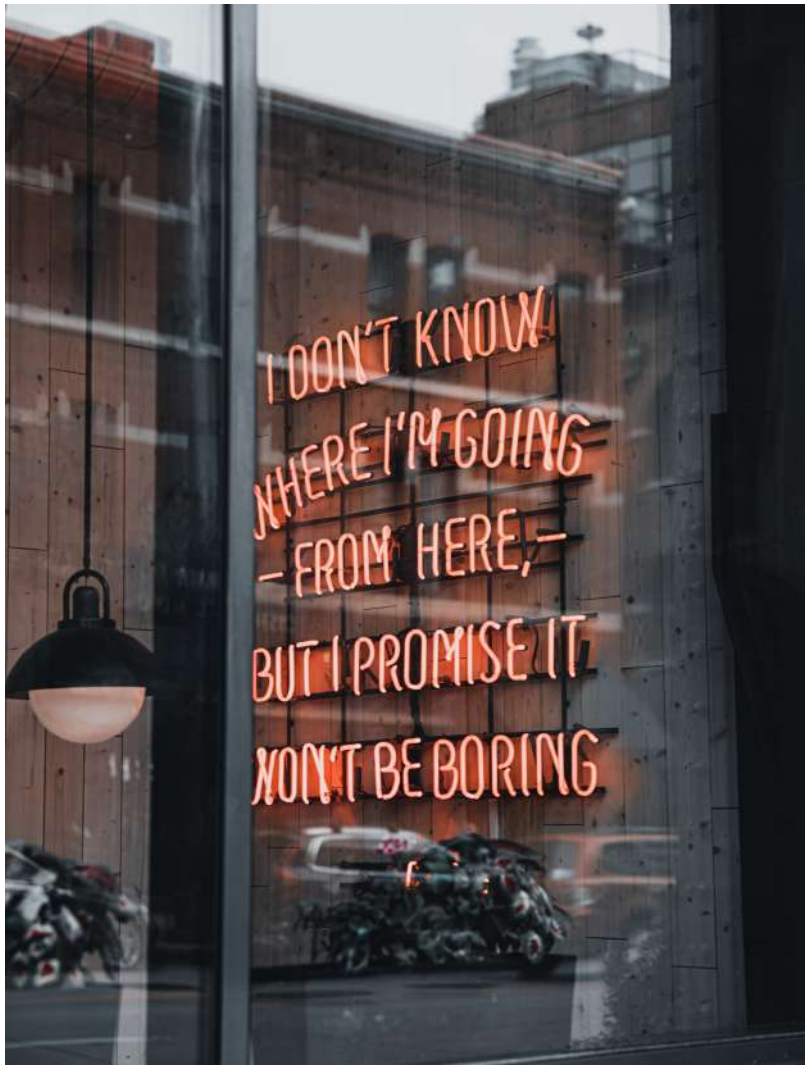
- Changing live through AI
- Passionate in robotics
- Data lover
- Films & series
- Bike enthusiast & martial artist practitioner

@ematde05

eduardo.matallanas@cabify.com

Things we will talk today

- How Cabify sees mobility
- Data Science in a changing environment
- Some ML stuff
- How the data is used to make decisions



Spoiler: The data journey starts here

cabify

Some of you will know about us…

Ride hailing app, competitors of

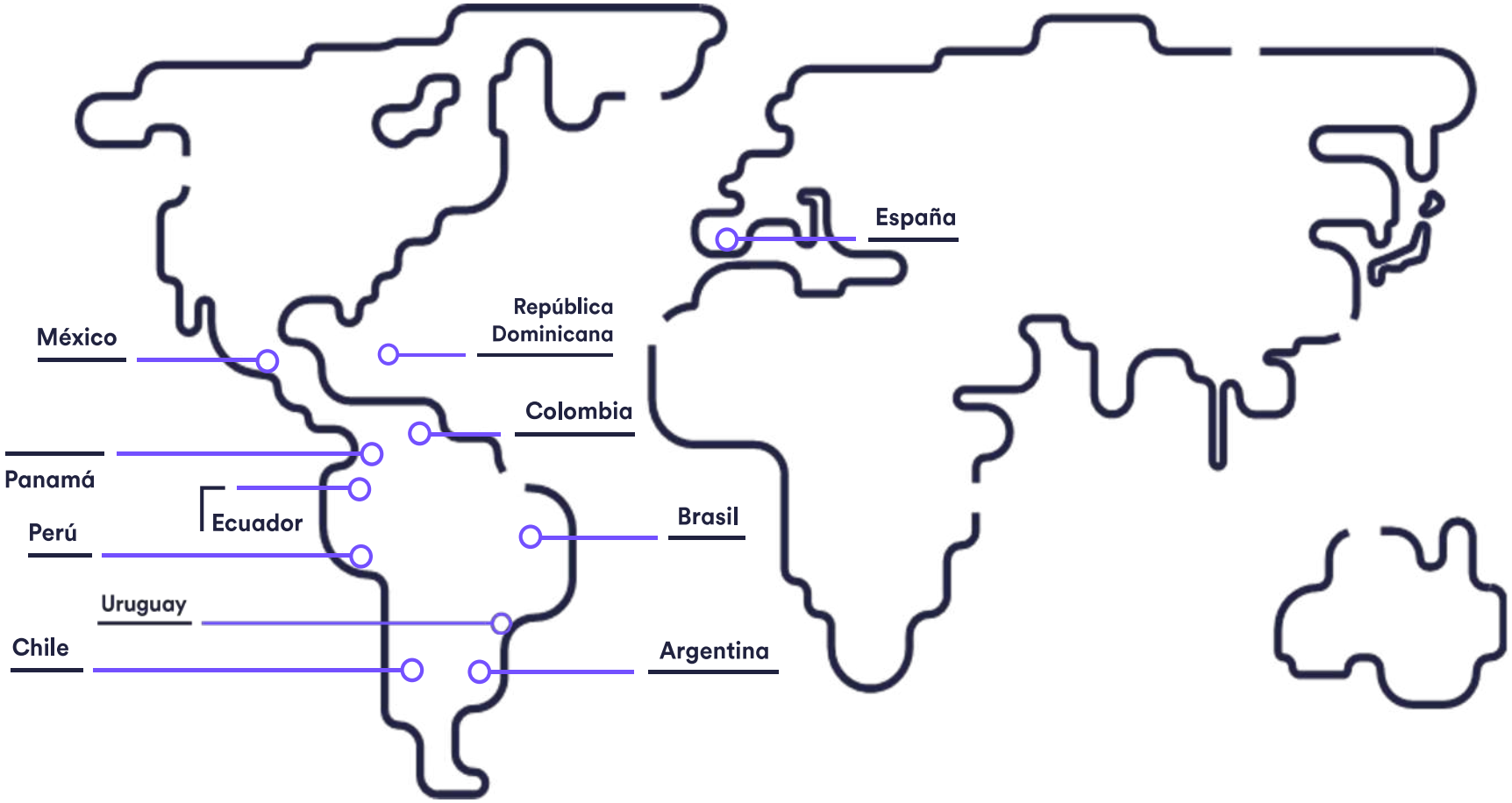
Uber

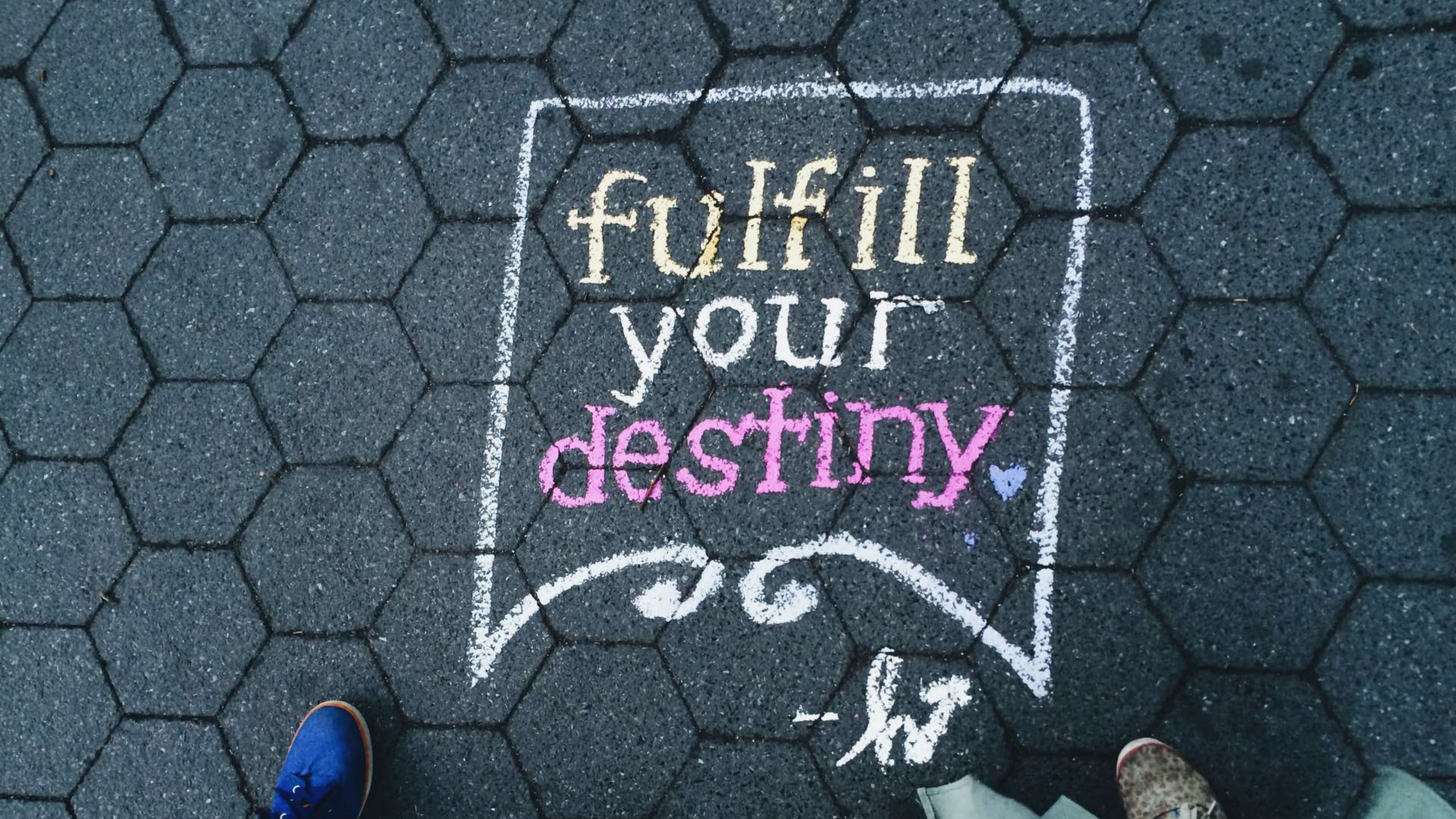
lyft

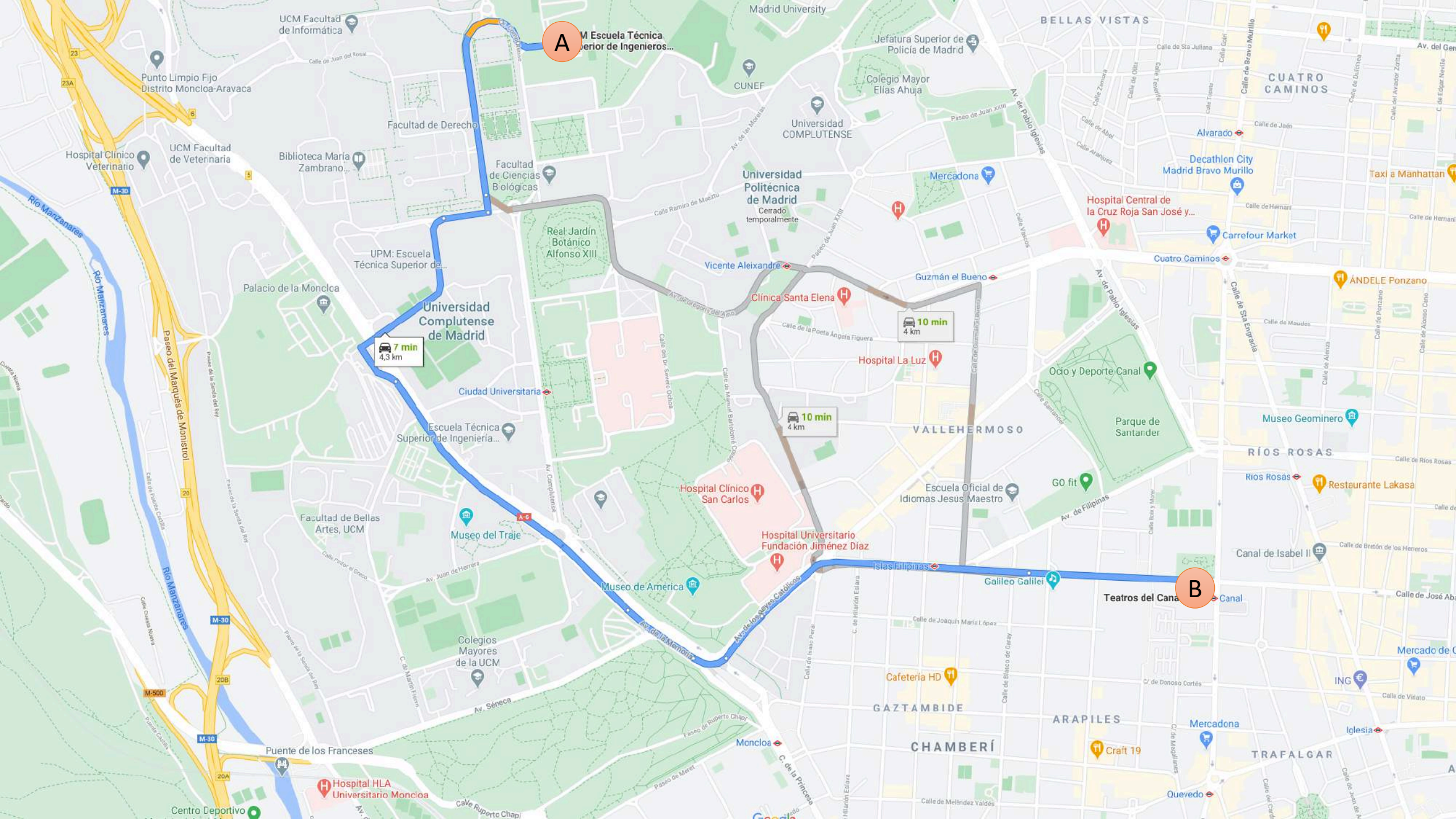
DiDi

The first Spanish Unicorn

We are present in



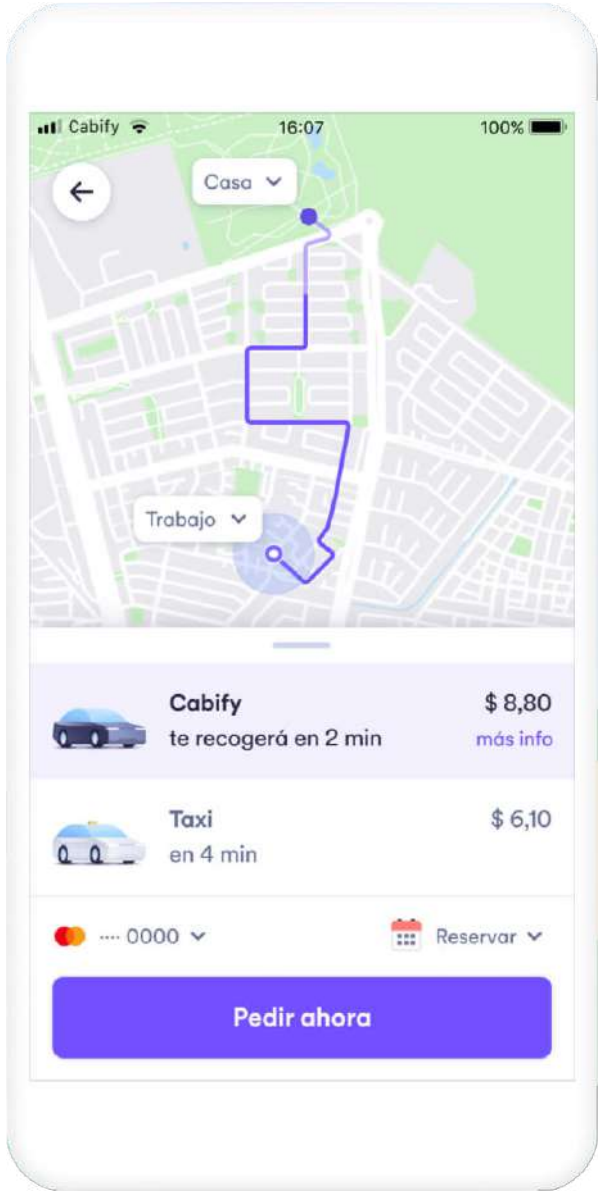

fulfill
your
destiny.


A
B
7 min
4,3 km
10 min
4 km
10 min
4 km
UCM Facultad de Informática
Punto Limpio Fijo Distrito Moncloa-Aravaca
Facultad de Derecho
Hospital Clínico Veterinario
UCM Facultad de Veterinaria
Biblioteca María Zambrano...
Facultad de Ciencias Biológicas
Universidad Politécnica de Madrid
Cerrado temporalmente
Real Jardín Botánico Alfonso XIII
UPM: Escuela Técnica Superior de...
Palacio de la Moncloa
Universidad Complutense de Madrid
Ciudad Universitaria
Escuela Técnica Superior de Ingeniería...
Facultad de Bellas Artes, UCM
Museo del Traje
Museo de América
Colegios Mayores de la UCM
Puente de los Franceses
Hospital HLA Universitario Moncloa
Centro Deportivo
Río Manzanares
Paseo del Marqués de Monistrol
Madrid University
CUNEF
Universidad COMPLUTENSE
Colegio Mayor Elías Ahuja
Jefatura Superior de Policía de Madrid
Mercadona
Vicente Aleixandre
Clínica Santa Elena
Hospital La Luz
Hospital Clínico San Carlos
Hospital Universitario Fundación Jiménez Díaz
Escuela Oficial de Idiomas Jesús Maestro
Guzmán el Bueno
Cuatro Caminos
Carrefour Market
Hospital Central de la Cruz Roja San José y...
Alvarado
Decathlon City Madrid Bravo Murillo
Taxi a Manhattan
ÁNDELE Ponzano
Ocio y Deporte Canal
Parque de Santander
Museo Geominero
Ríos Rosas
Restaurante Lakasa
GO fit
Canal de Isabel II
Islas Filipinas
Galileo Galilei
Teatros del Canal
Cafetería HD
Craft 19
Mercadona
ING
Iglesia
Quevedo
Moncloa
BELLAS VISTAS
CUATRO CAMINOS
VALLEHERMOSO
RÍOS ROSAS
GAZTAMBIDE
ARAPILES
CHAMBERÍ
TRAFALGAR

SINGULARITY
TECH DAY

cabify
#STechDay2020

Looking for a ride?



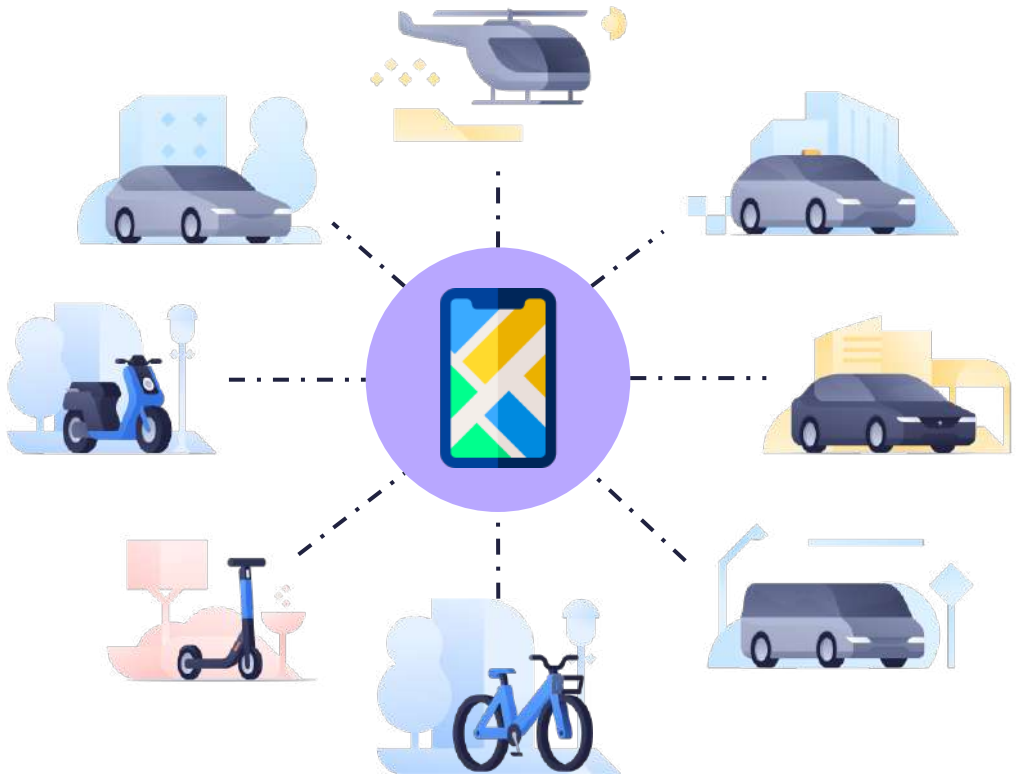
It's a Match!

We are not only a ride, we are a journey

- Safe and reliable
- Alternative to private transportation
- Transforming cities
- Being sustainable
- Embrace the change

In MaaS we trust

MaaS i.e. Mobility as a Service

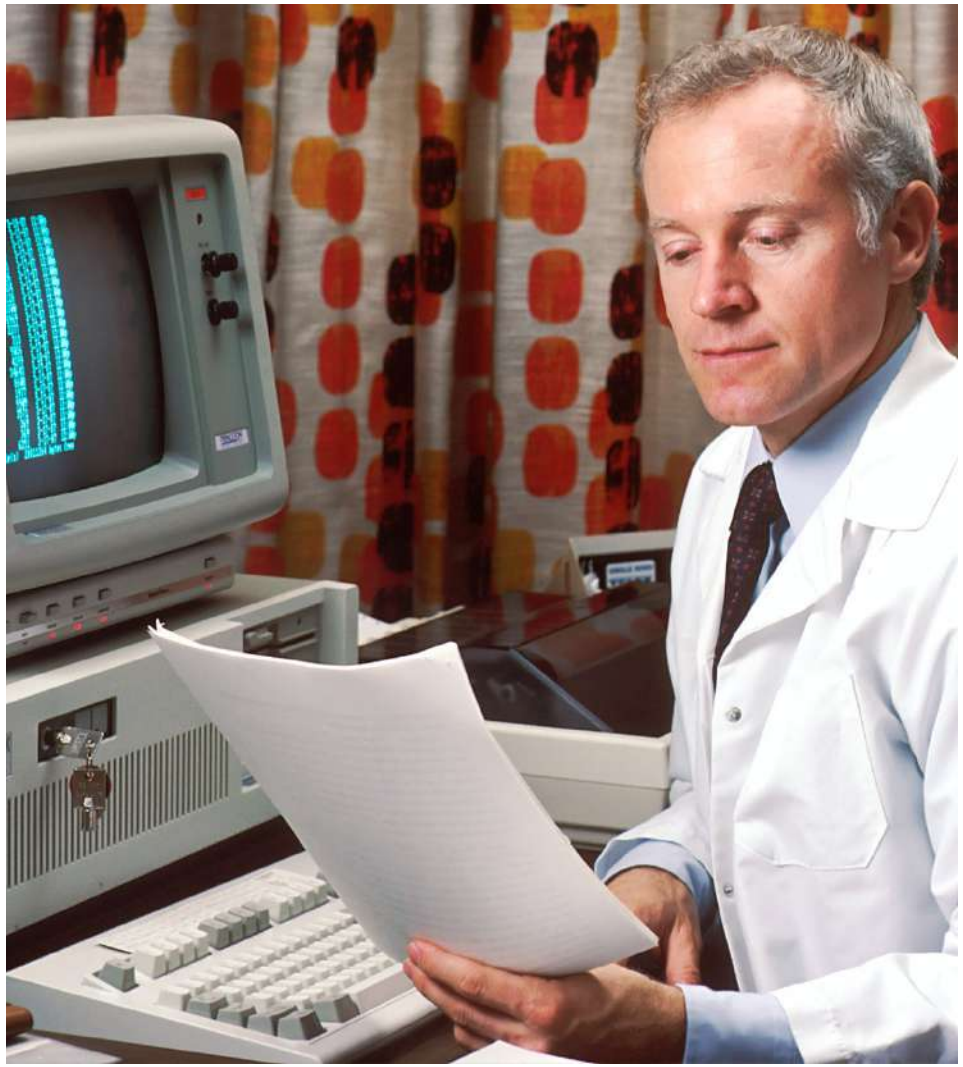


All these journeys are stored and transformed into wisdom

Who create knowledge?

- The Data Science
  - Turn harvested data into knowledge
  - Decisions made on facts not in evidence
  - Test your hypothesis

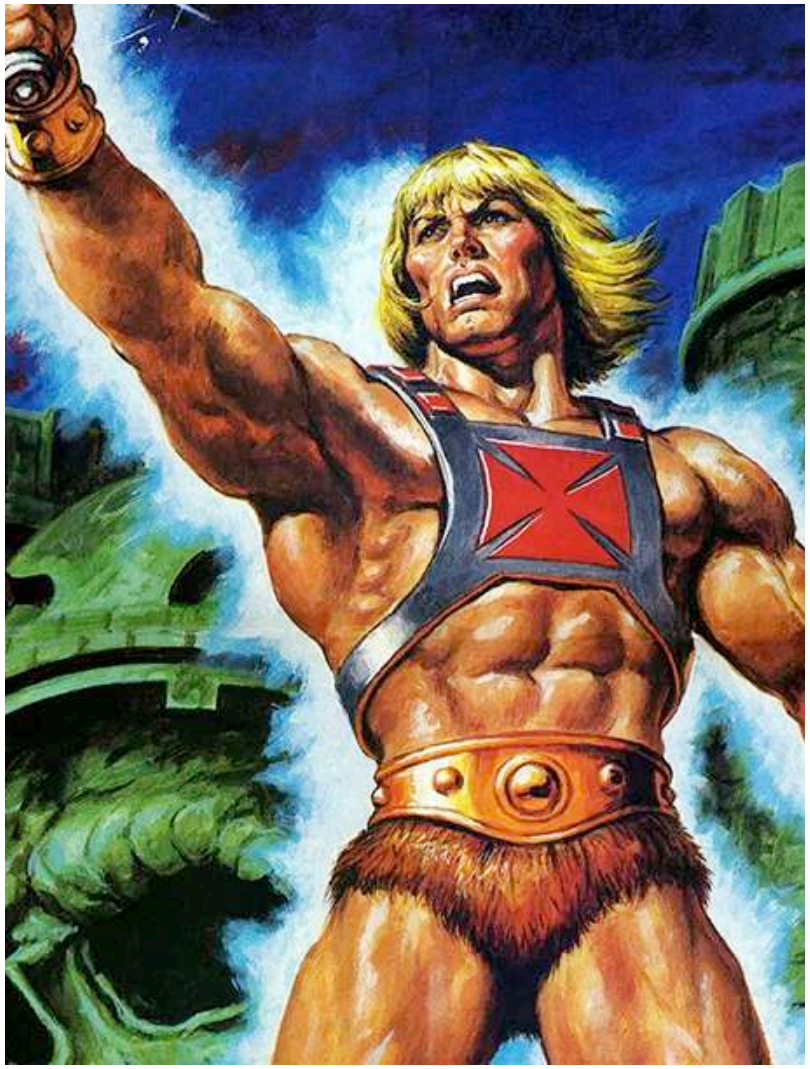
- Results are translated into
  - Better experiences for users
  - Well balanced marketplace
  - Improving our cities, making more sustainable

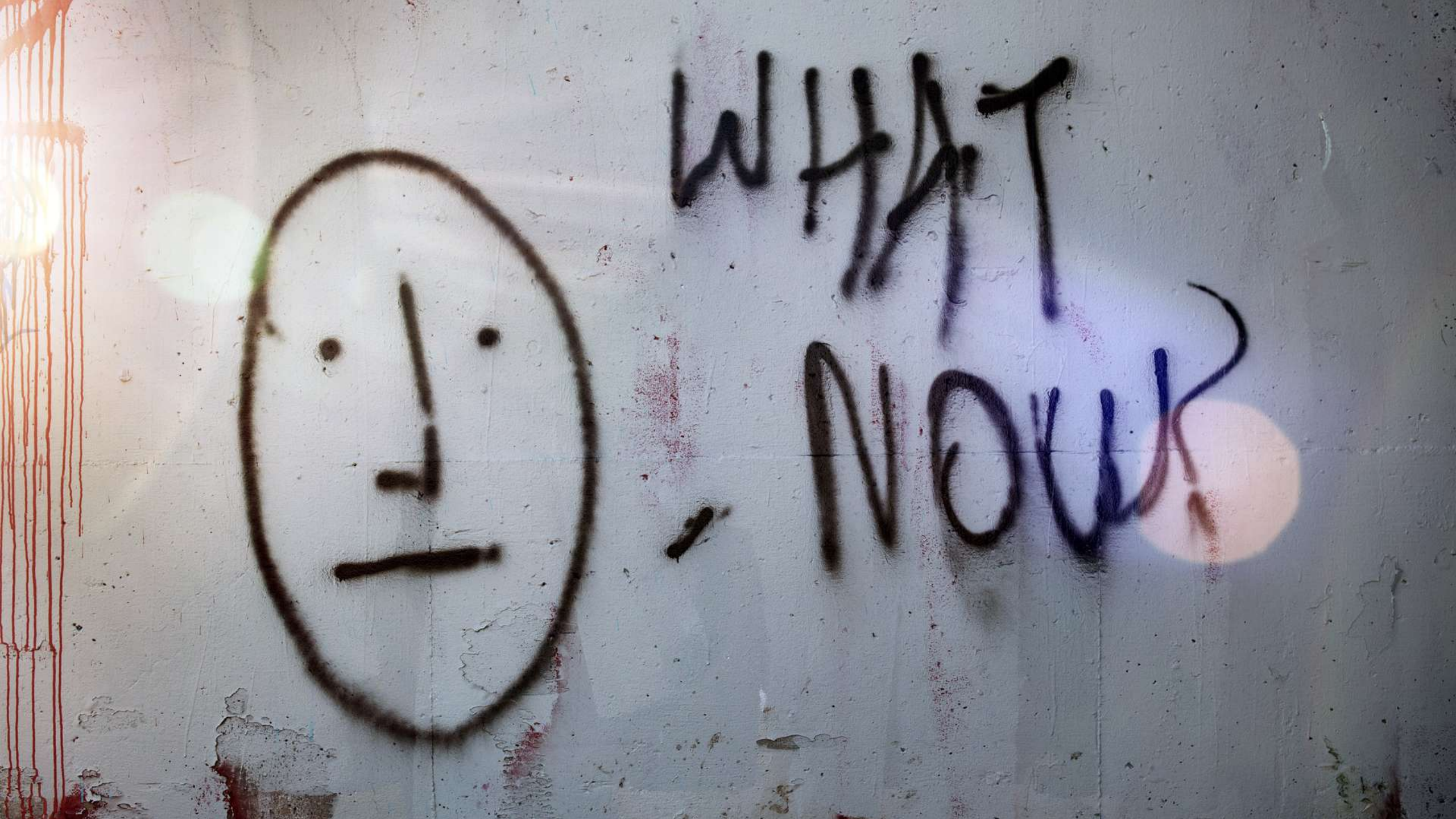


And how are the results leveraged?

- We have the Research Link
  - Close to other squads and managers
  - See opportunities through science
  - Act as a supporter with facts
  - Close the gap with engineering
  - Take decisions to the end

Make Knowledge Great Again




WHAT
NOW?

Stages of a DS project

Rigorously estimate size of the problem

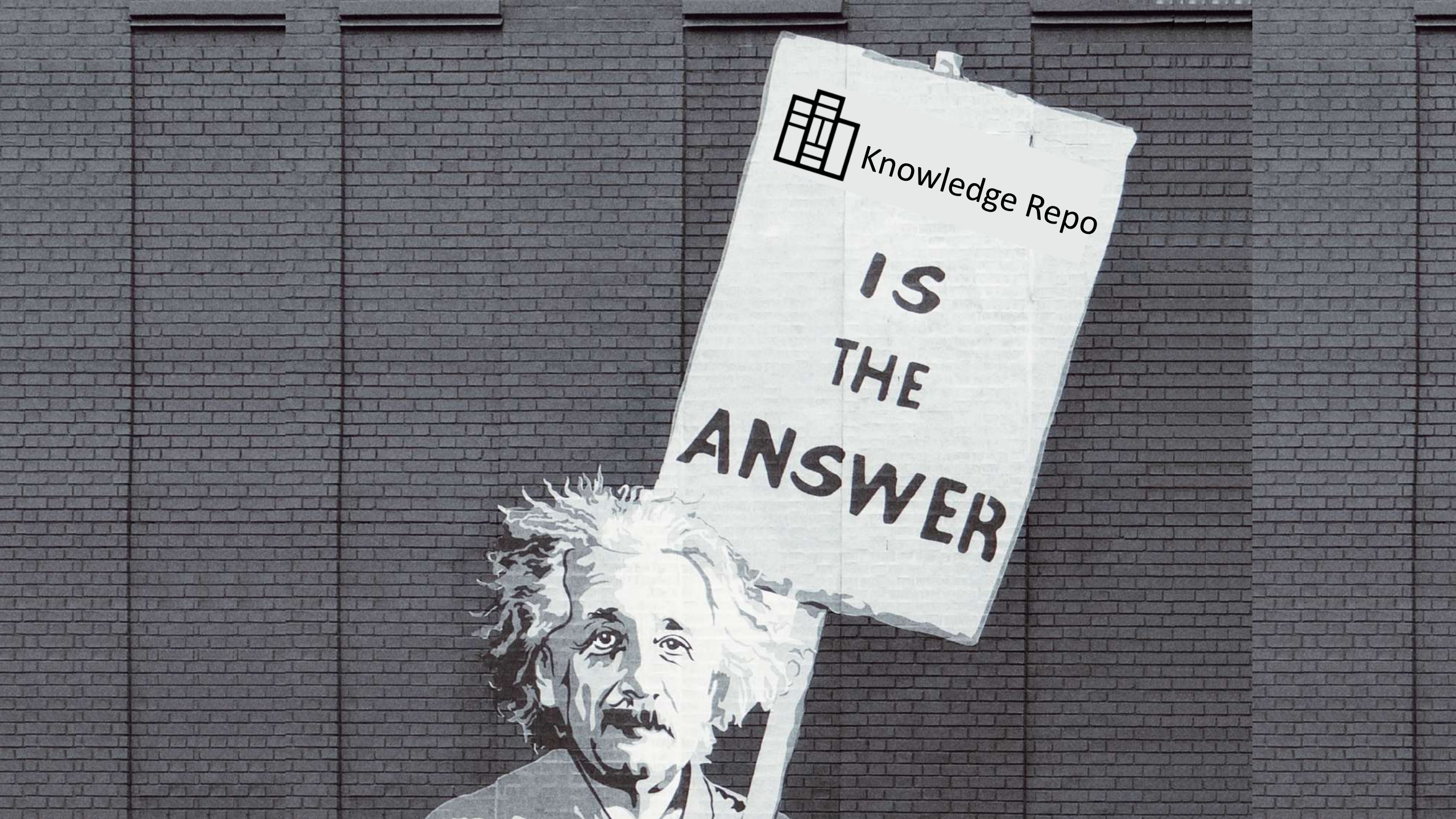
Prototype working with cold data or in dark launch

Deploy the model to affect the user experience

Measure the impact, establishing relationships after the release

Scaling Knowledge @ Cabify

- Must guarantee:
  - Quality
  - Reproducibility
  - Accessibility
  - Discoverability


Knowledge Repo
IS
THE
ANSWER

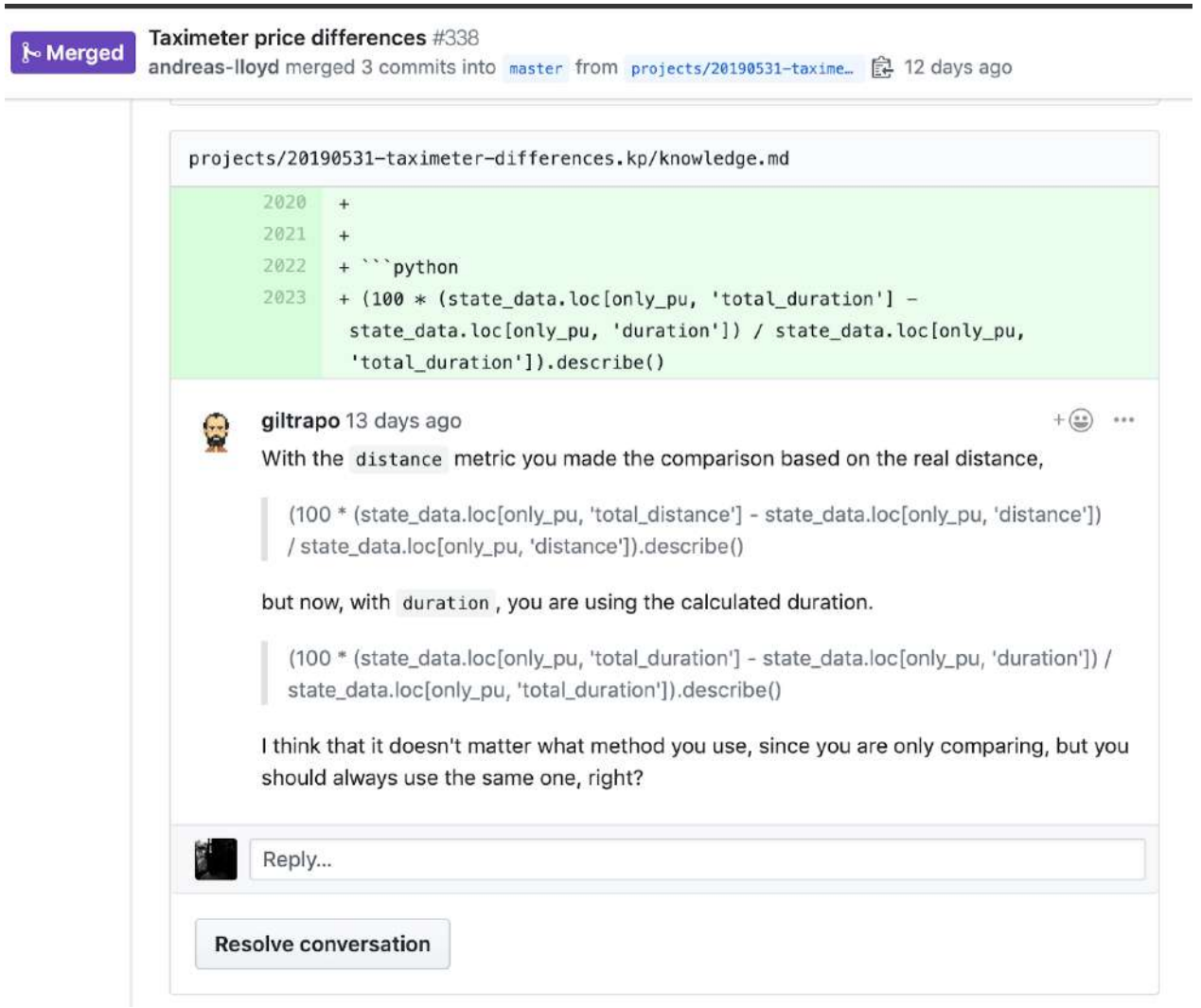
Quality

Through revision process

non-reviewable.ipynb

↓

reviewable.md



Reproducibility

- Control-versioned notebooks code
- Frozen data
- Track of project advance
- Environment & libraries involved

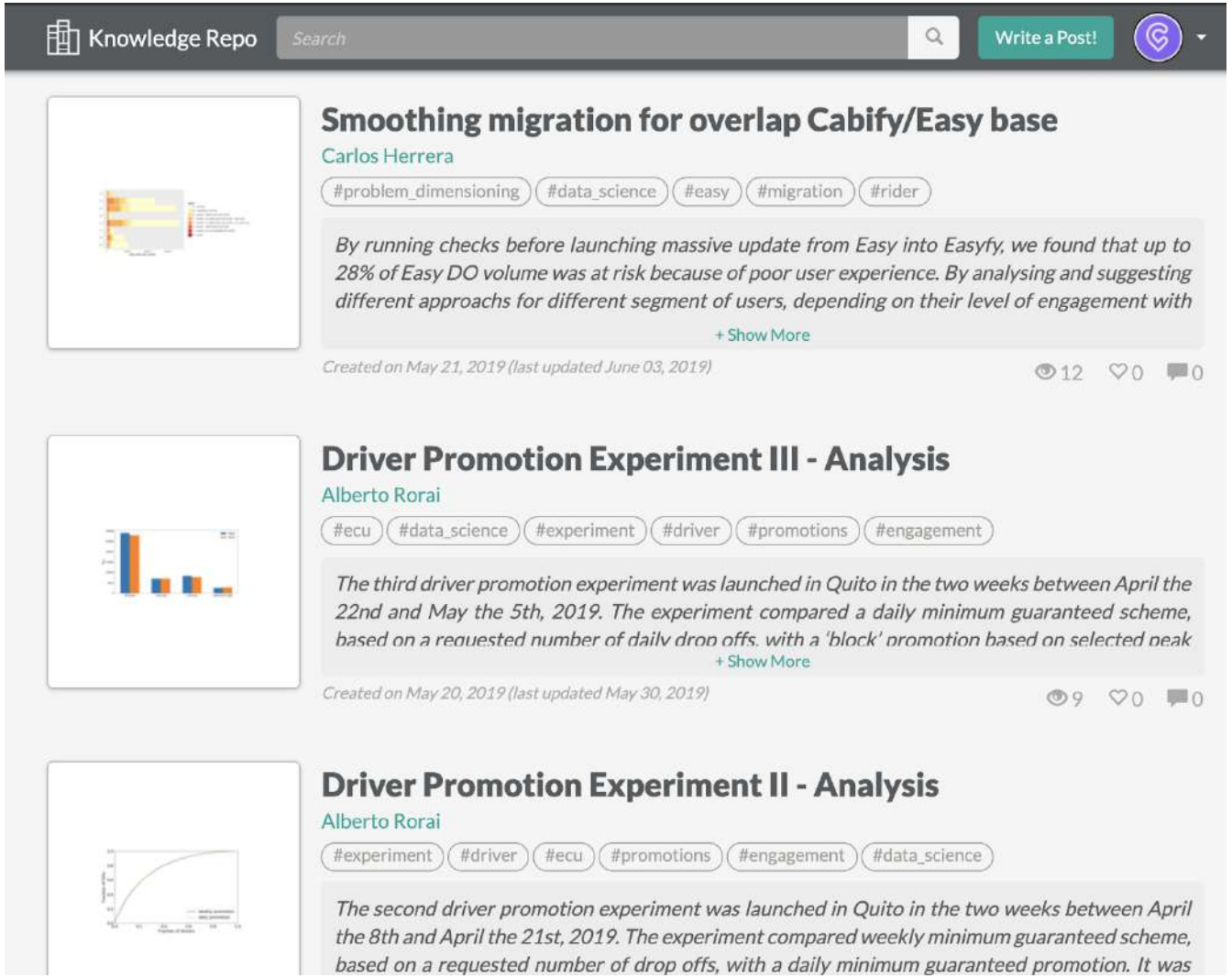
```
get_requirements(globals())

cabify-data-utils==0.3.2
matplotlib==3.1.2
numpy==1.17.4
pandas==1.0.5
plotly==4.8.2
plotnine==0.7.0
science-tools==0.1a0
seaborn==0.9.0
statsmodels==0.11.1
```

- Complete repository: cabiMaps
- Data files: available at cabify-research-experiments/cabiMaps/
  - 1.4K files, 9 folder, 645.9 GiB total size
- Clubhouse stories: 82006

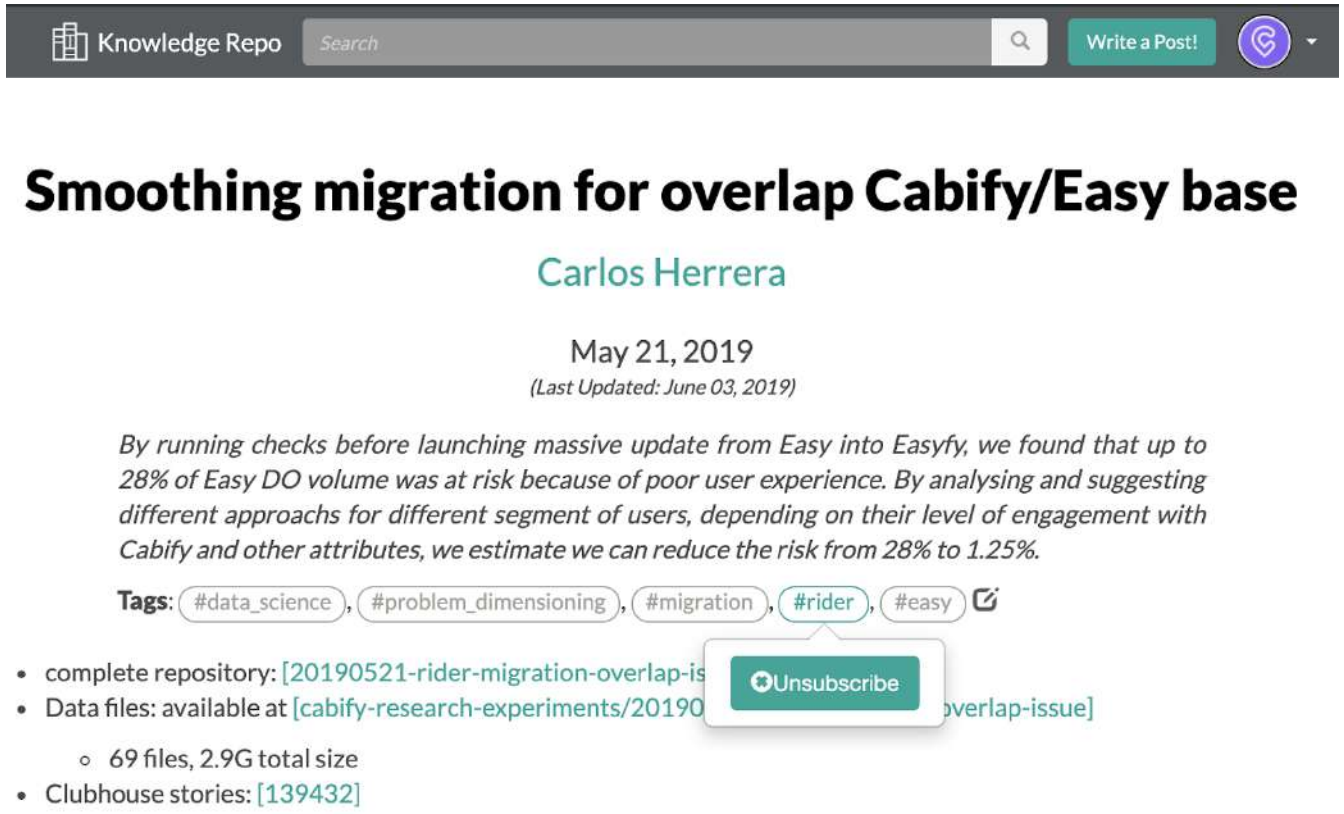
Accesibility

- TLDR + Key Figure
- Brief executive summary
  - Context & Motivation
  - What we did
  - Key Figure
  - Conclusions & recommendations
- Large Scientific report
  - Contains all insights
  - The necessary code



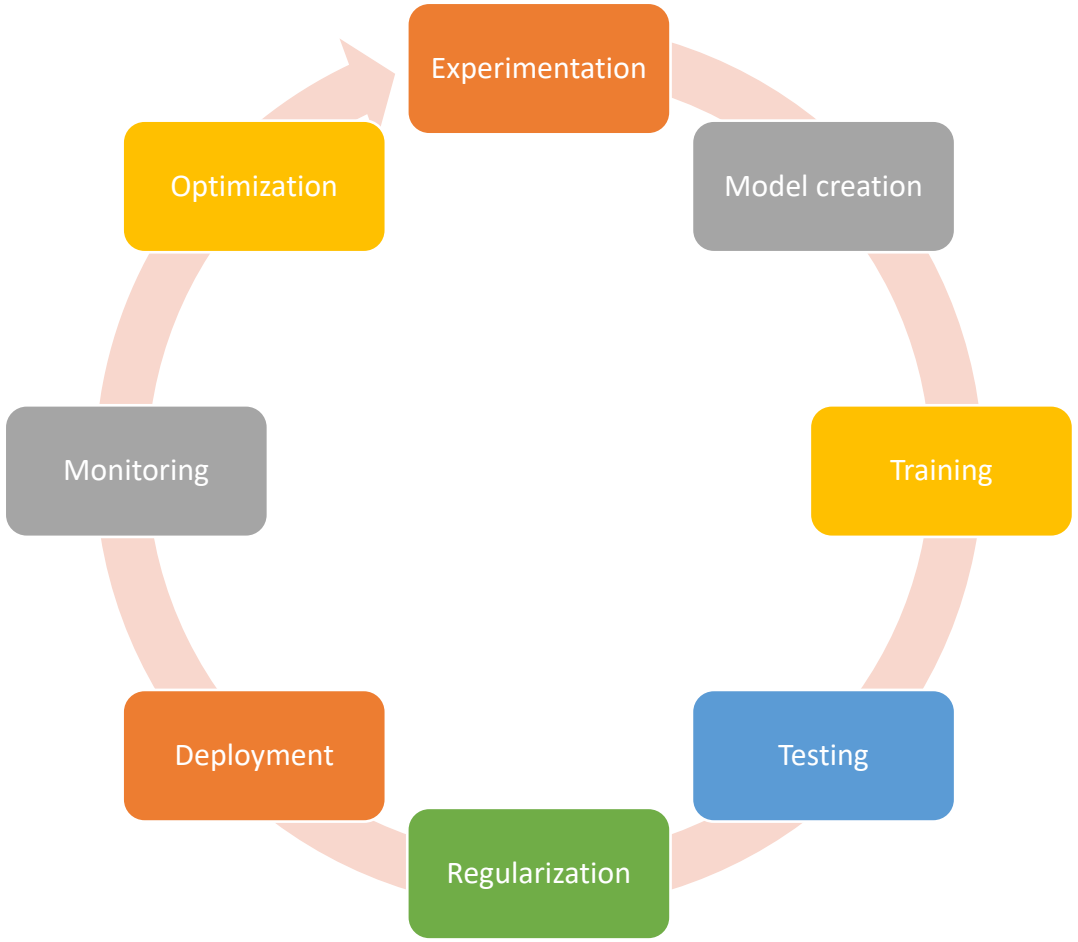
Discoverability

- Search bar
- Tags
- Subscription



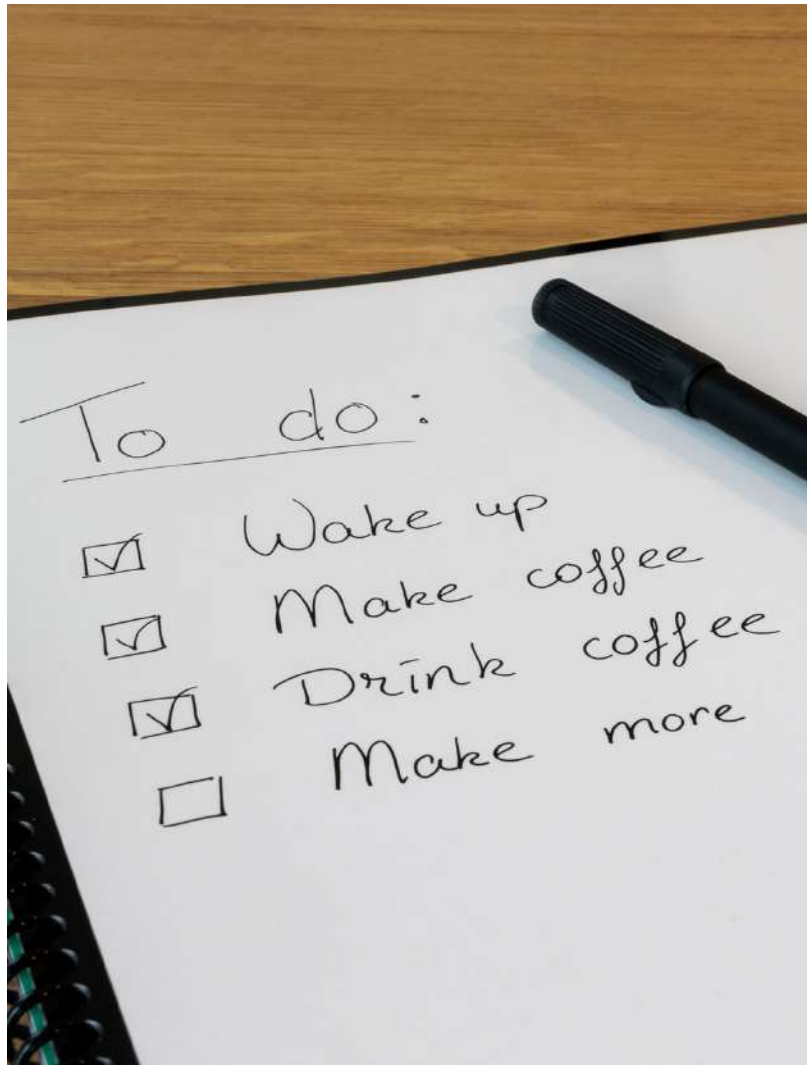
Stages of a DS project

Machine Learning Lifecycle



Wish list

- Train and evaluate their models
- Put models in production alone
- Fix the gap between training and deployment
- Monitor model performance
- Ensure reproducibility of our models



Unleash experimentation

Cabify’s machine learning platform to rule them all

What is Lykeion?

- The solution to all ML production ready results
  - Features → ”Scientific” data built to help
  - Models → Learn from data, need to be inferred
- Based on:
  - GCP → BigQuery, Kubernetes, etc.
  - Mlflow → Model serving heart
  - Prometheus → Monitoring
  - Grafana → Visualizing, monitoring and alert
- We are refining the monitoring part of the platform

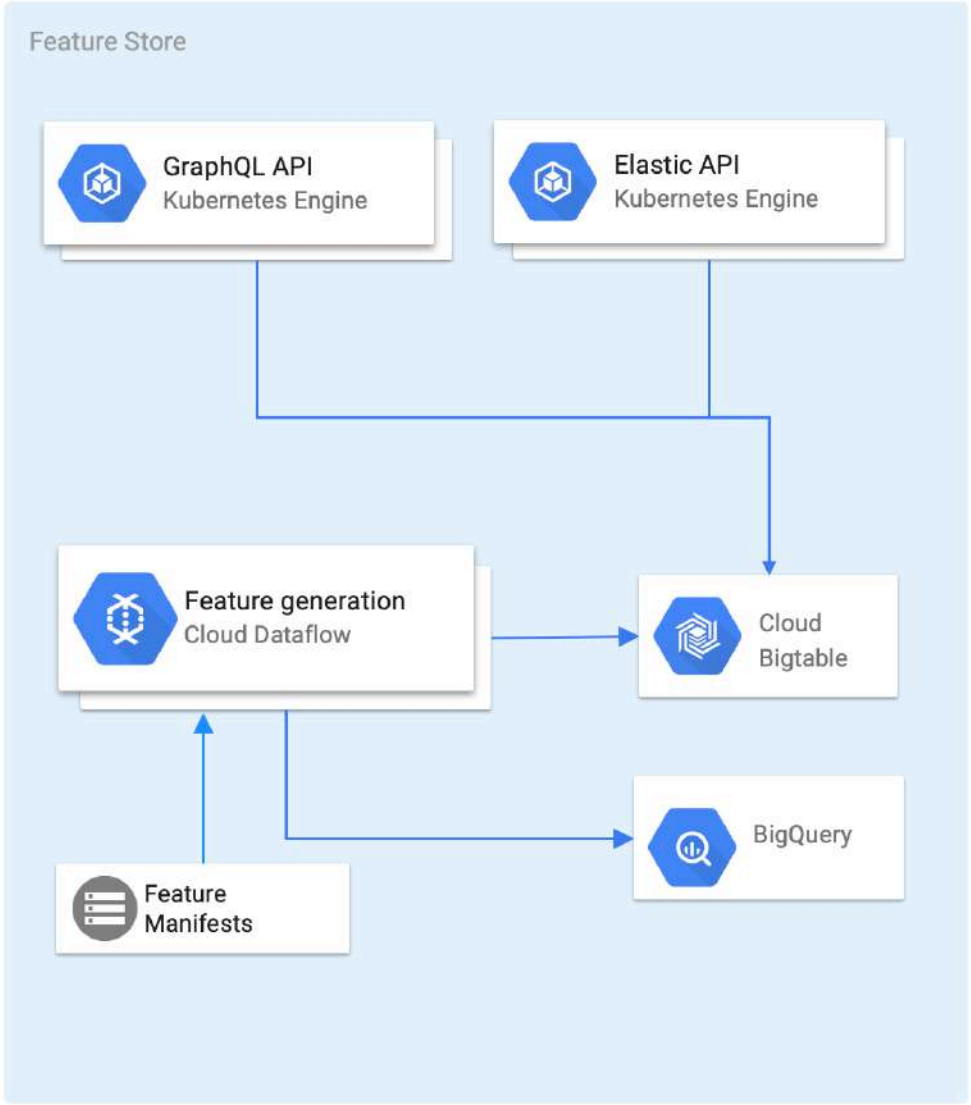
Features, the great unknown

Theory → individual measurable property or characteristic of a phenomenon being observed

Practice → augmented data that allows to take decisions from business

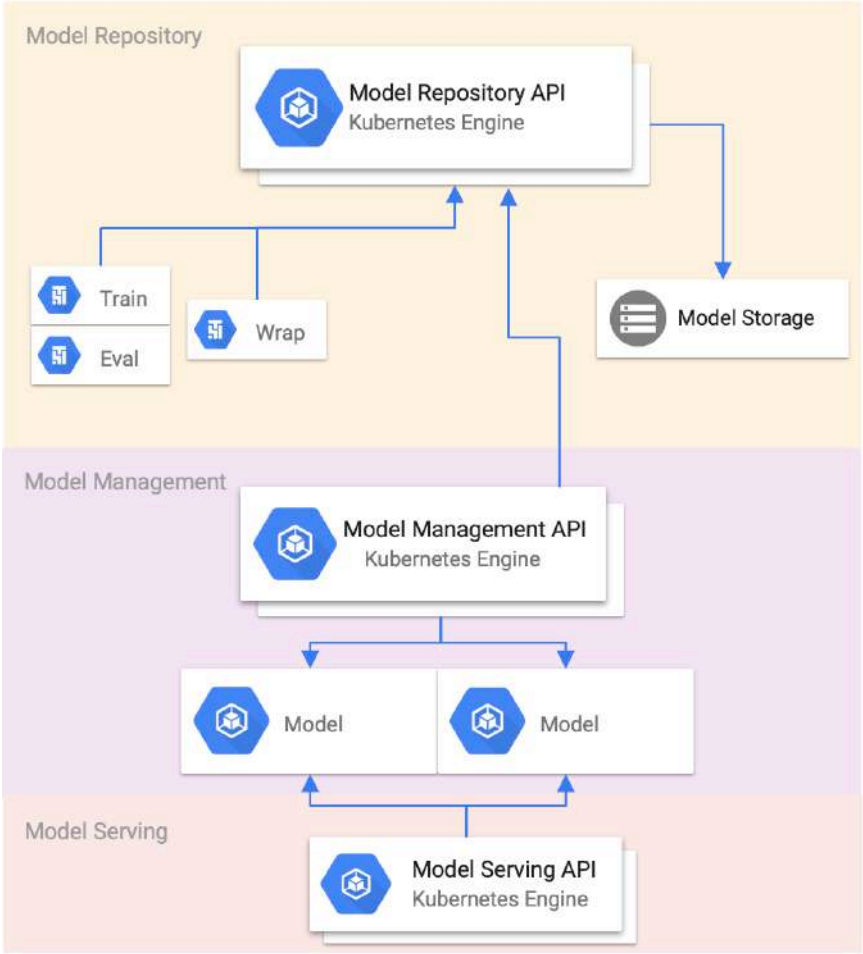
Feature Store

- Add more value in less time
- Reuse features calculated by others
  - Models
  - Business
- Different storages for different purposes
  - Big Table serves them
  - BigQuery generate them and train
- Consistent way of calculating features

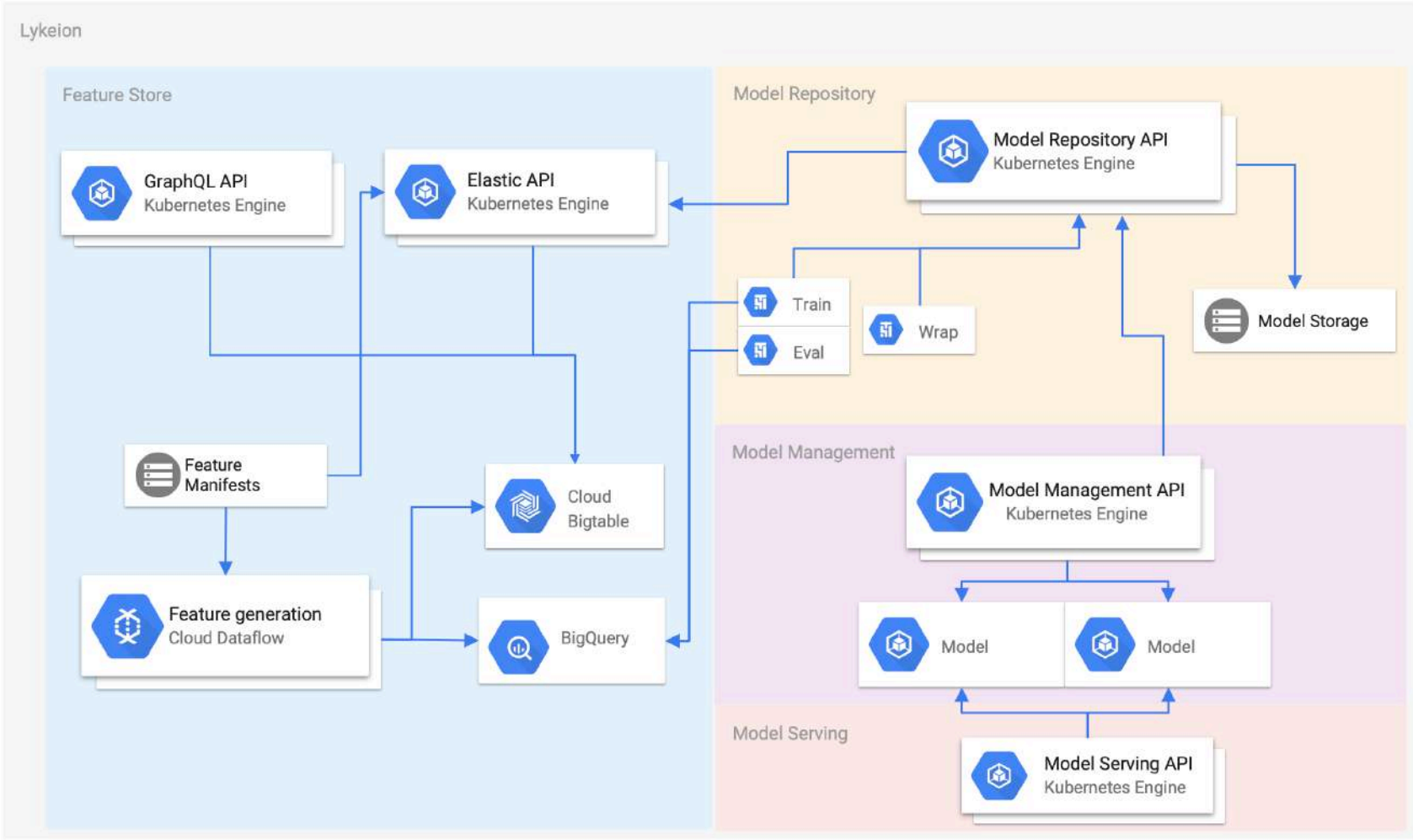


ML Serving platform

1. Model Repository, based on MLflow
   - Prototype your model
   - Use features
   - Store resultant model
2. Model Management API
   - Handle stored models
   - Deploy model version in prod
3. Model Serving API
   - Exposed the model to the world
   - Access control
   - Monitoring models

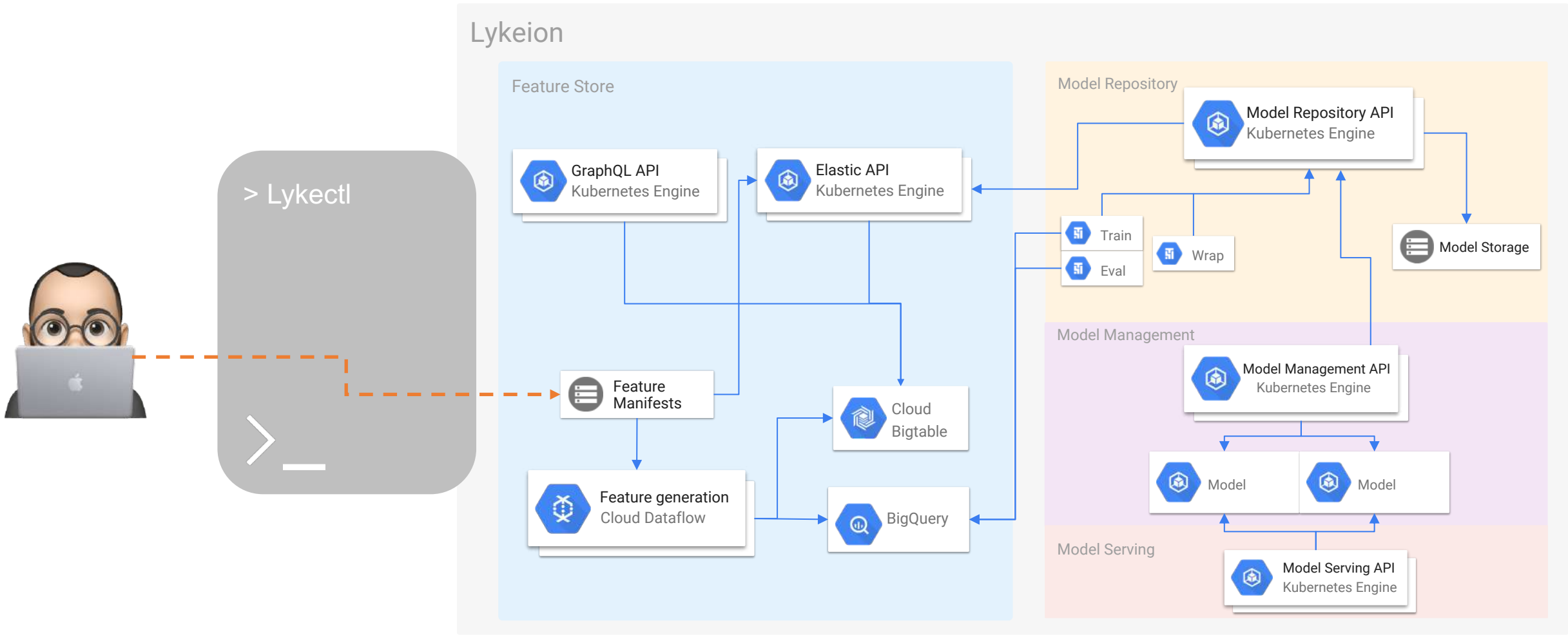


The whole picture

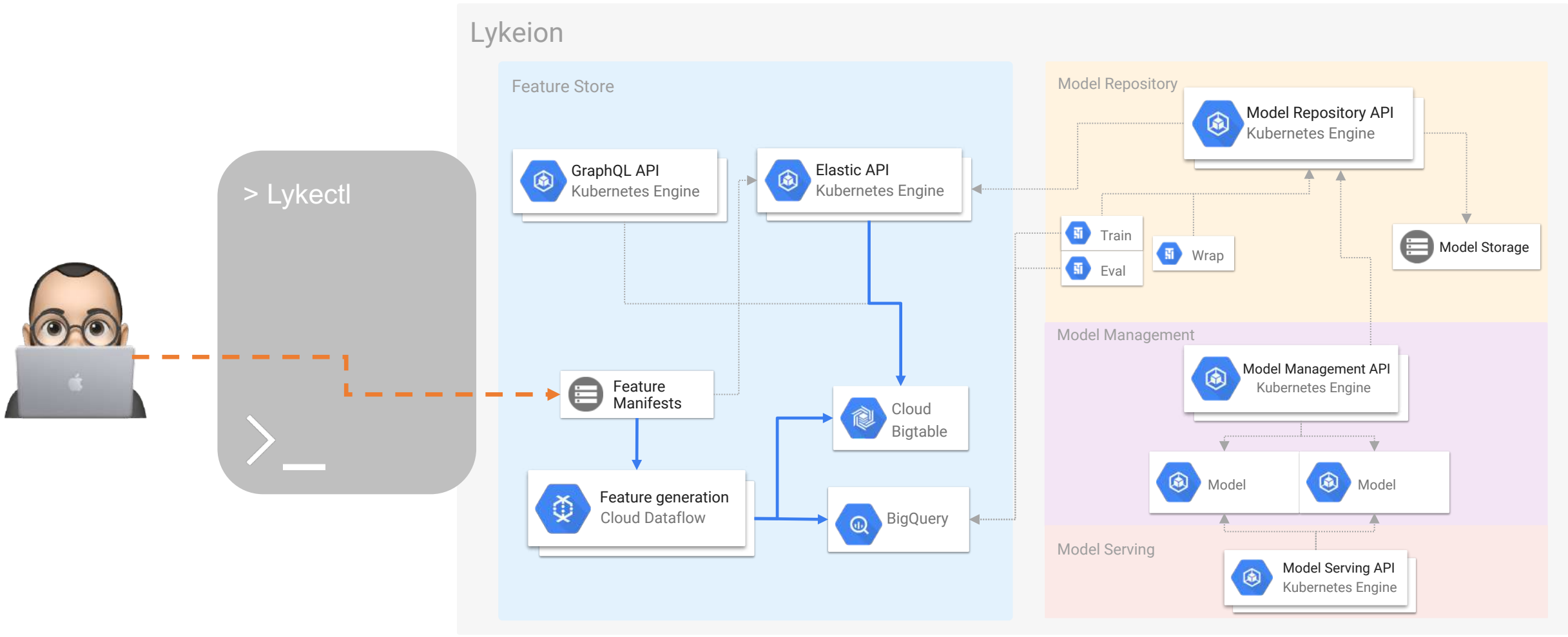


From zero to production

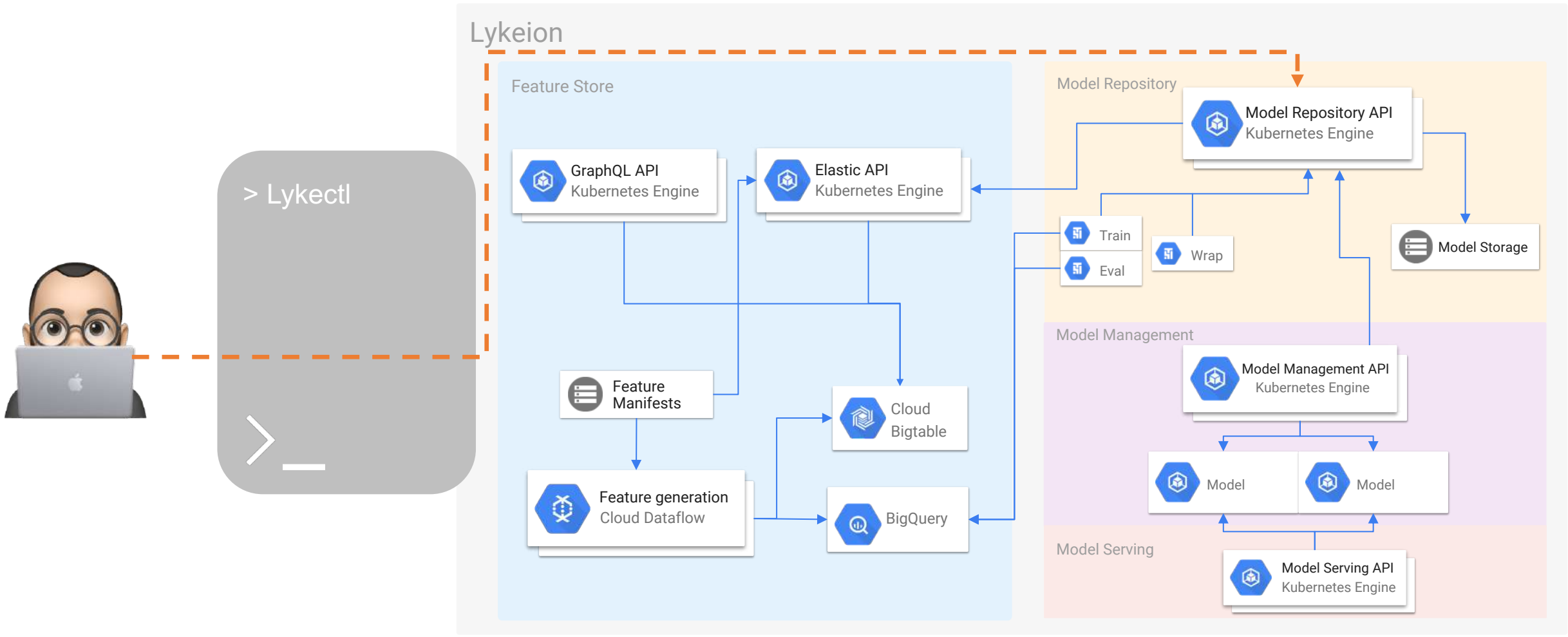
How easy is to create a feature?



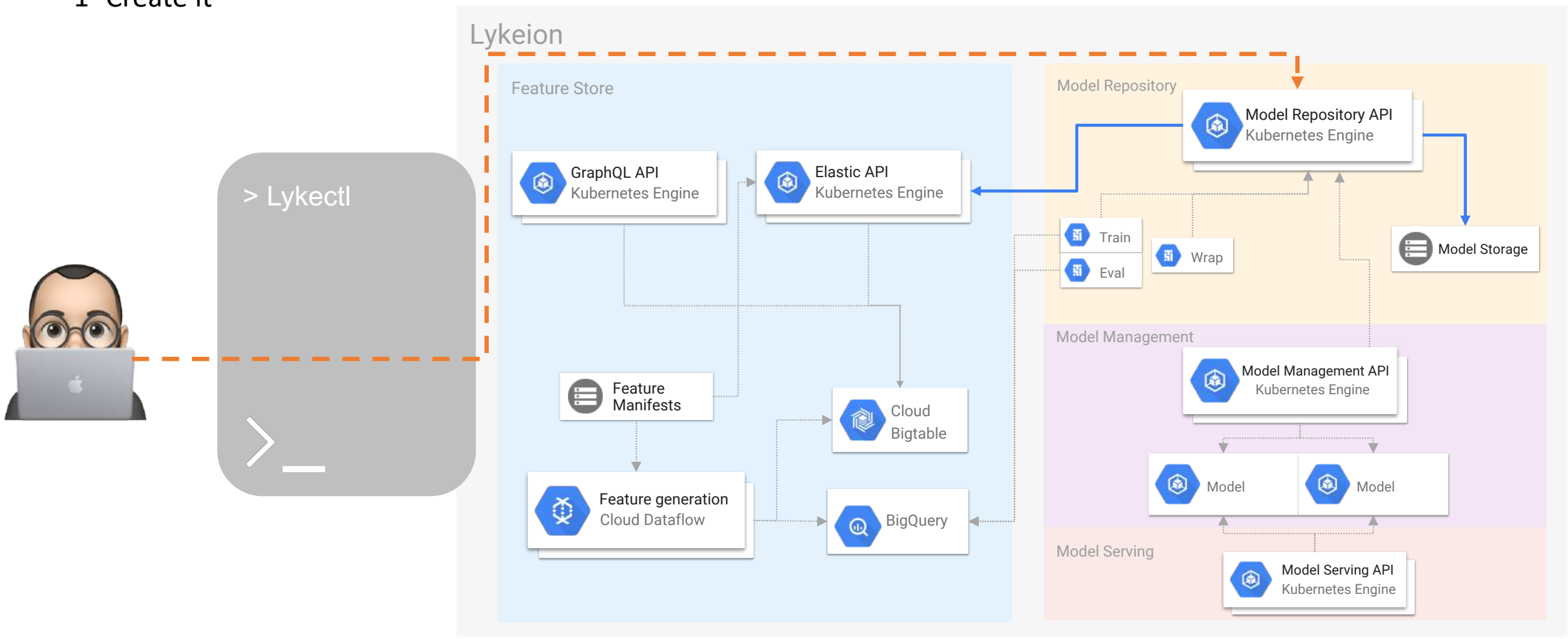
How easy is to create a feature?



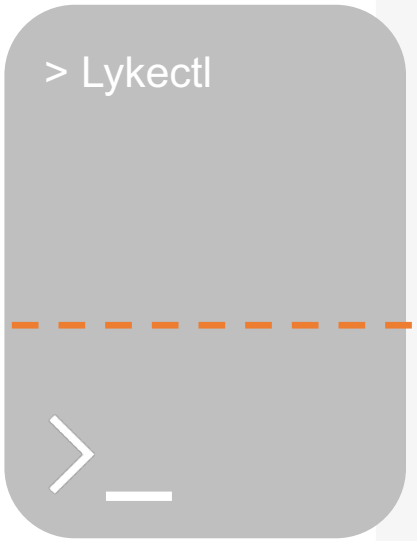
And what about model prototyping?

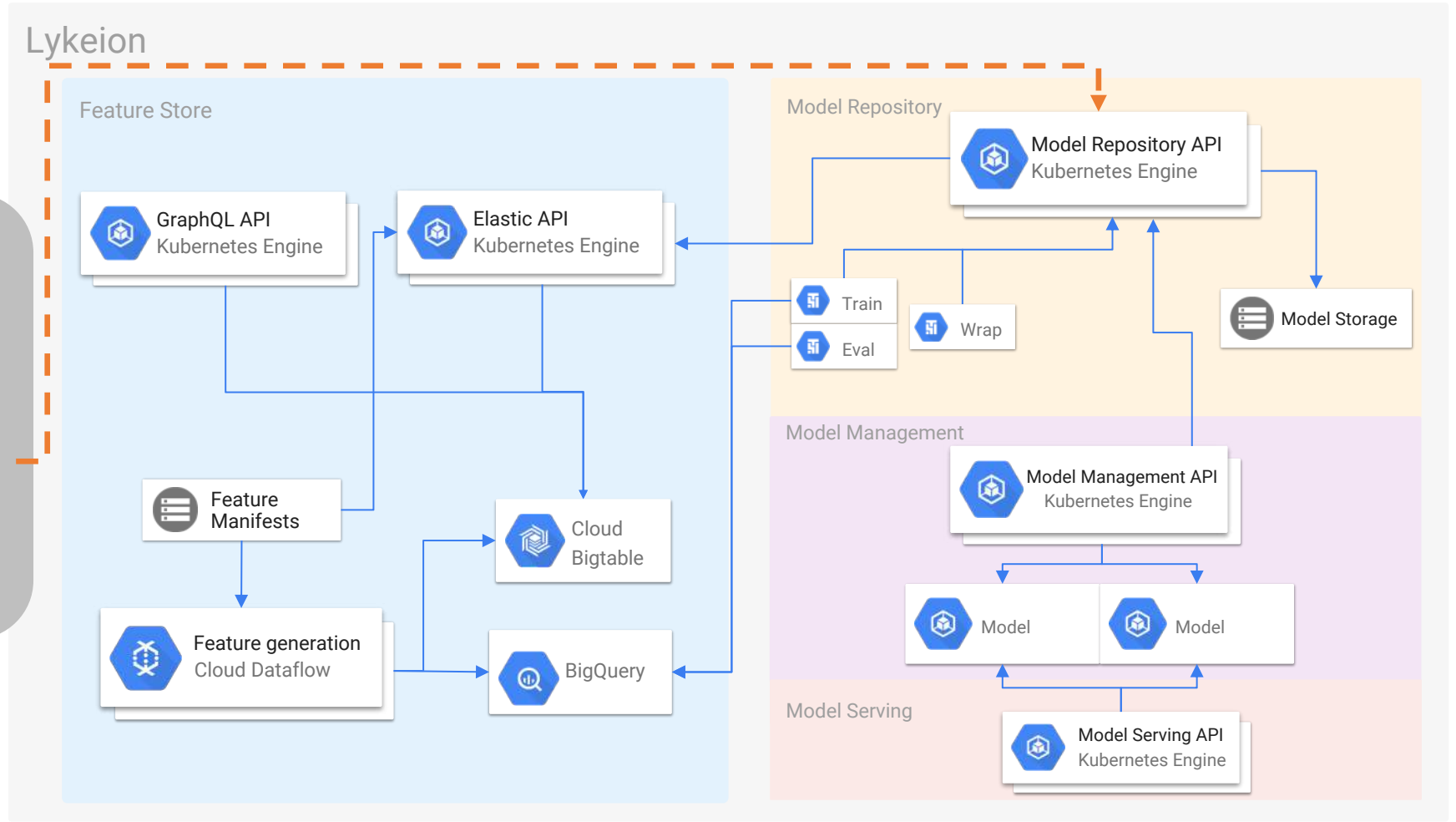


And what about model prototyping?
1- Create it

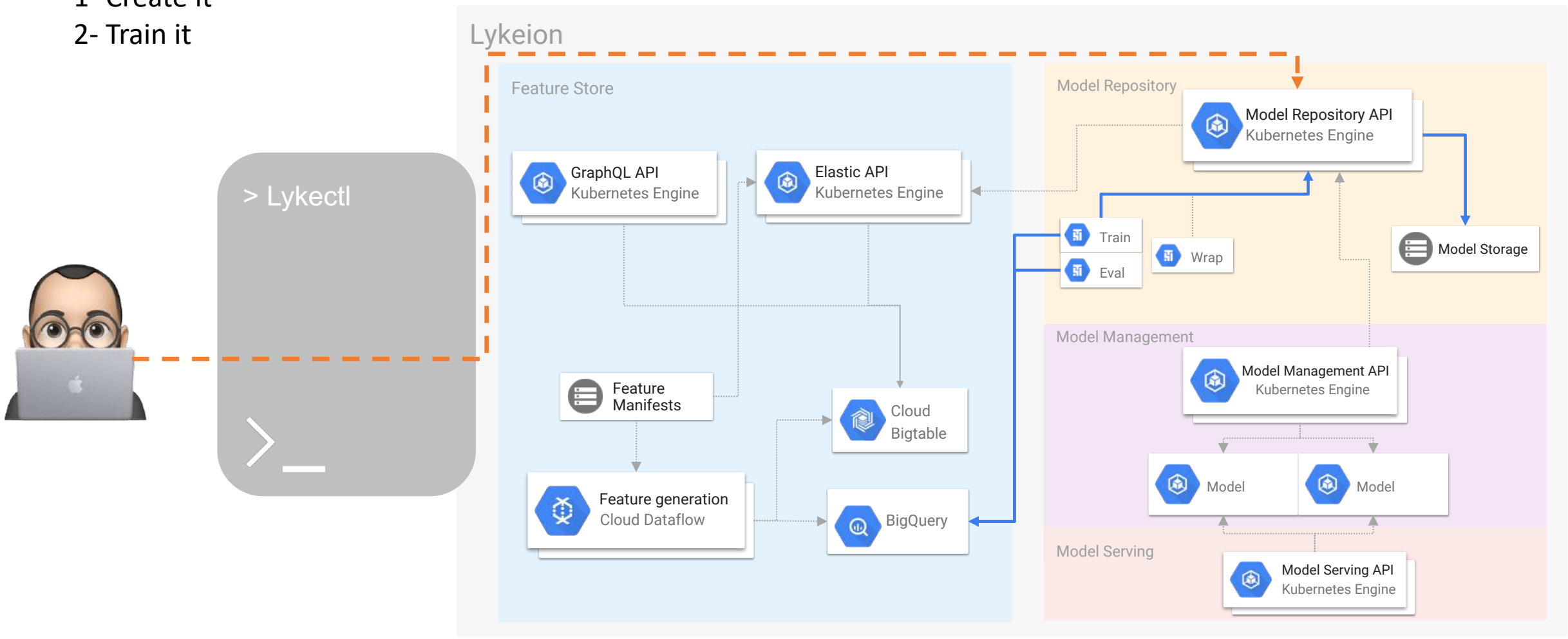


And what about model prototyping?
1- Create it
2- Train it

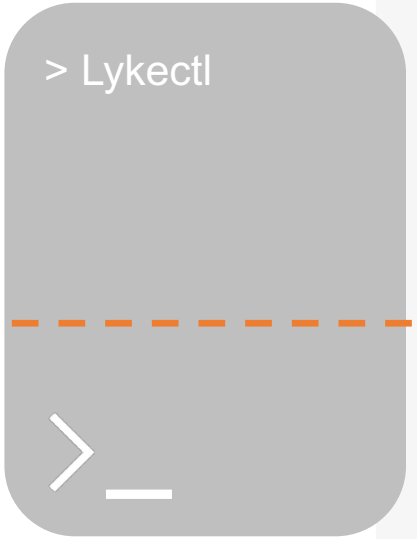


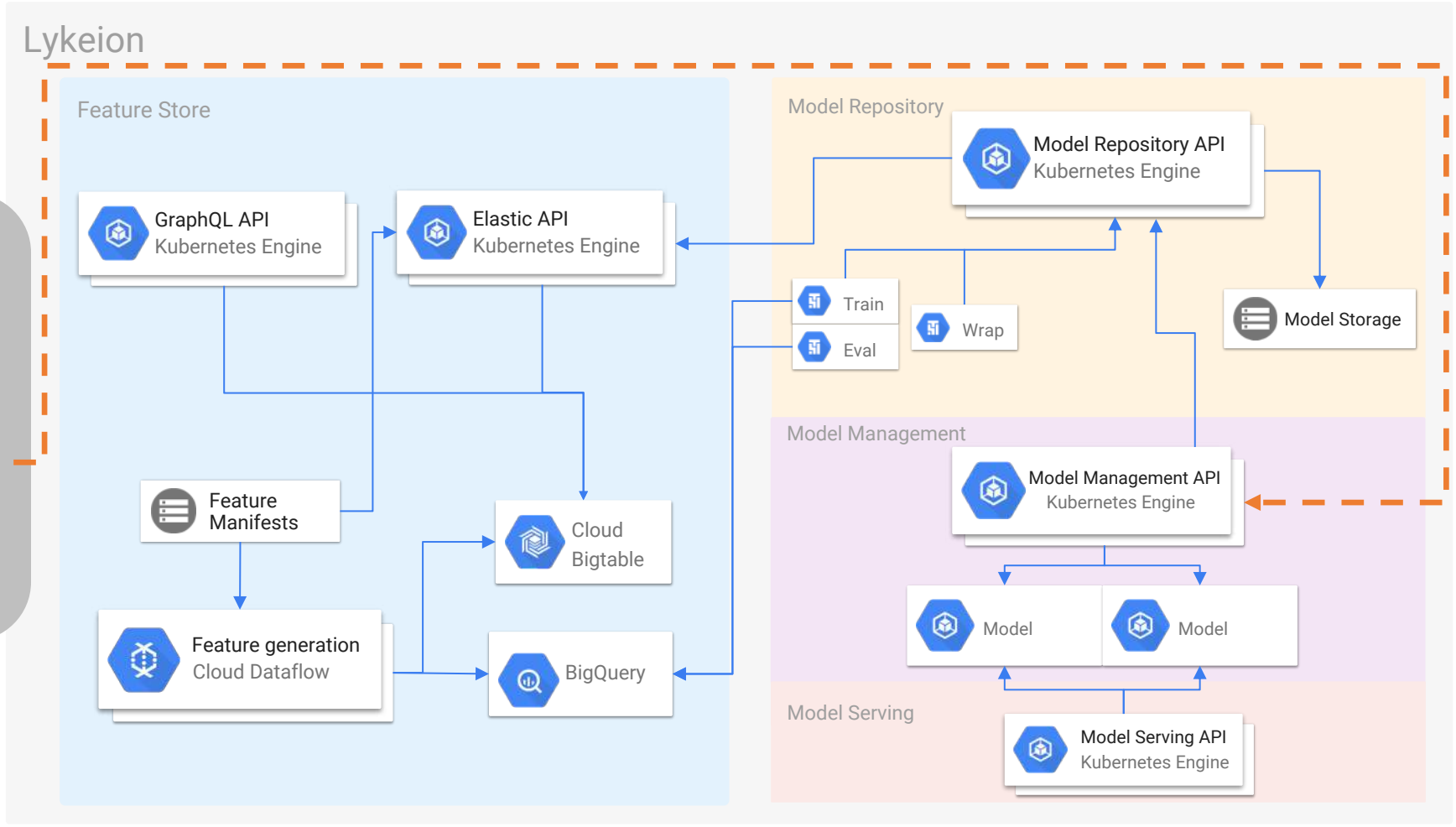


And what about model prototyping?
1- Create it
2- Train it

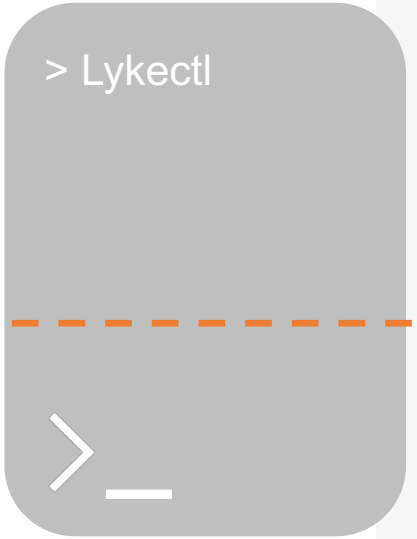


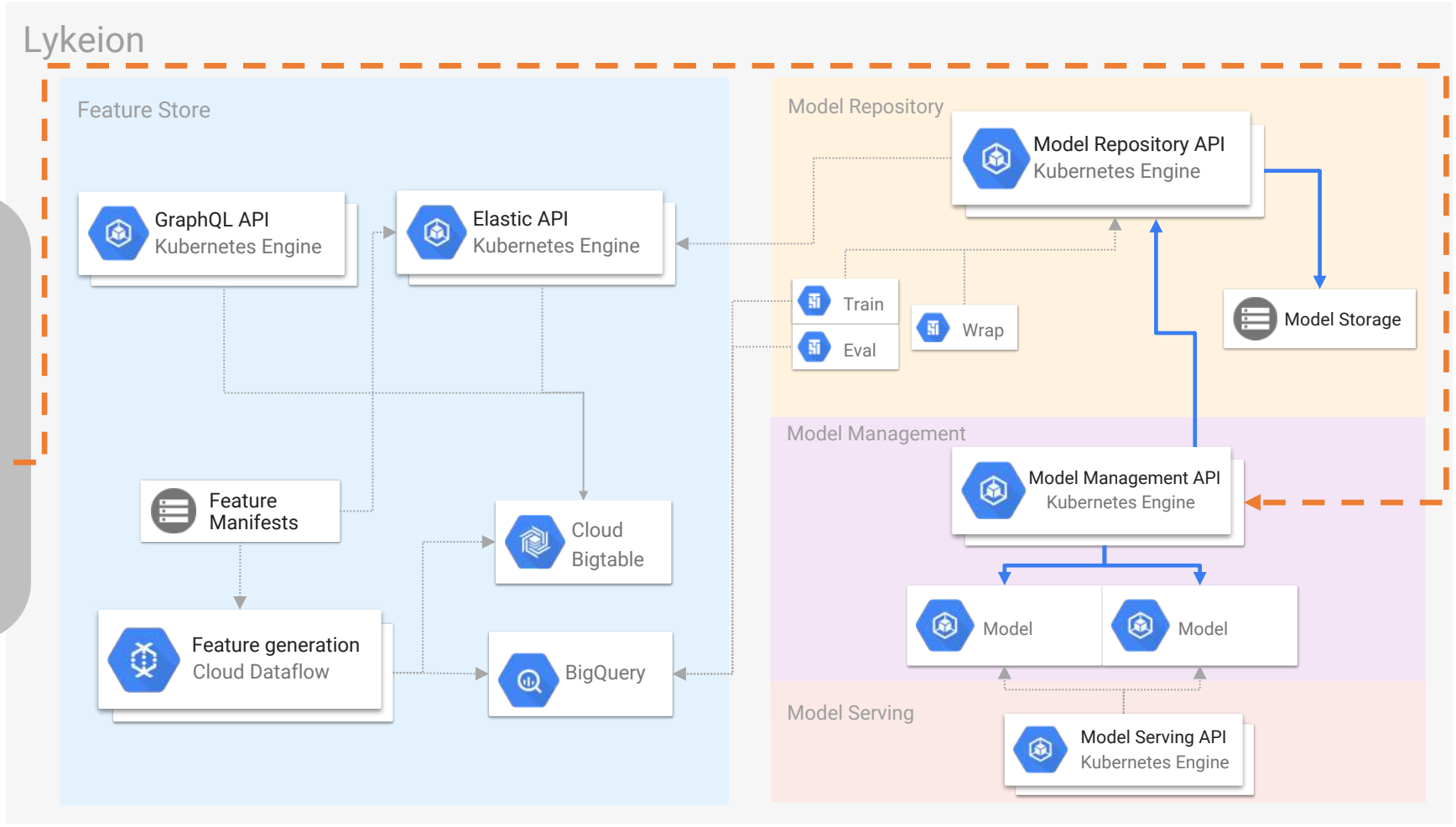
And what about model prototyping?
1- Create it
2- Train it
3- Deploy it





And what about model prototyping?
1- Create it
2- Train it
3- Deploy it





And what about model prototyping?

1- Create it
2- Train it
3- Deploy it
4- Infer it

> Lykectl

Lykeion

Feature Store

GraphQL API
Kubernetes Engine

Elastic API
Kubernetes Engine

Feature
Manifests

Cloud
Bigtable

Feature generation
Cloud Dataflow

BigQuery

Model Repository

Model Repository API
Kubernetes Engine

Train

Eval

Wrap

Model Storage

Model Management

Model Management API
Kubernetes Engine

Model

Model

Model Serving

Model Serving API
Kubernetes Engine

SINGULARITY
TECH DAY

cabify
#STechDay2020

And what about model prototyping?

1- Create it
2- Train it
3- Deploy it
4- Infer it

> Lykectl

Lykeion

Feature Store

GraphQL API
Kubernetes Engine

Elastic API
Kubernetes Engine

Feature Manifests

Cloud Bigtable

Feature generation
Cloud Dataflow

BigQuery

Model Repository

Model Repository API
Kubernetes Engine

Train

Eval

Wrap

Model Storage

Model Management

Model Management API
Kubernetes Engine

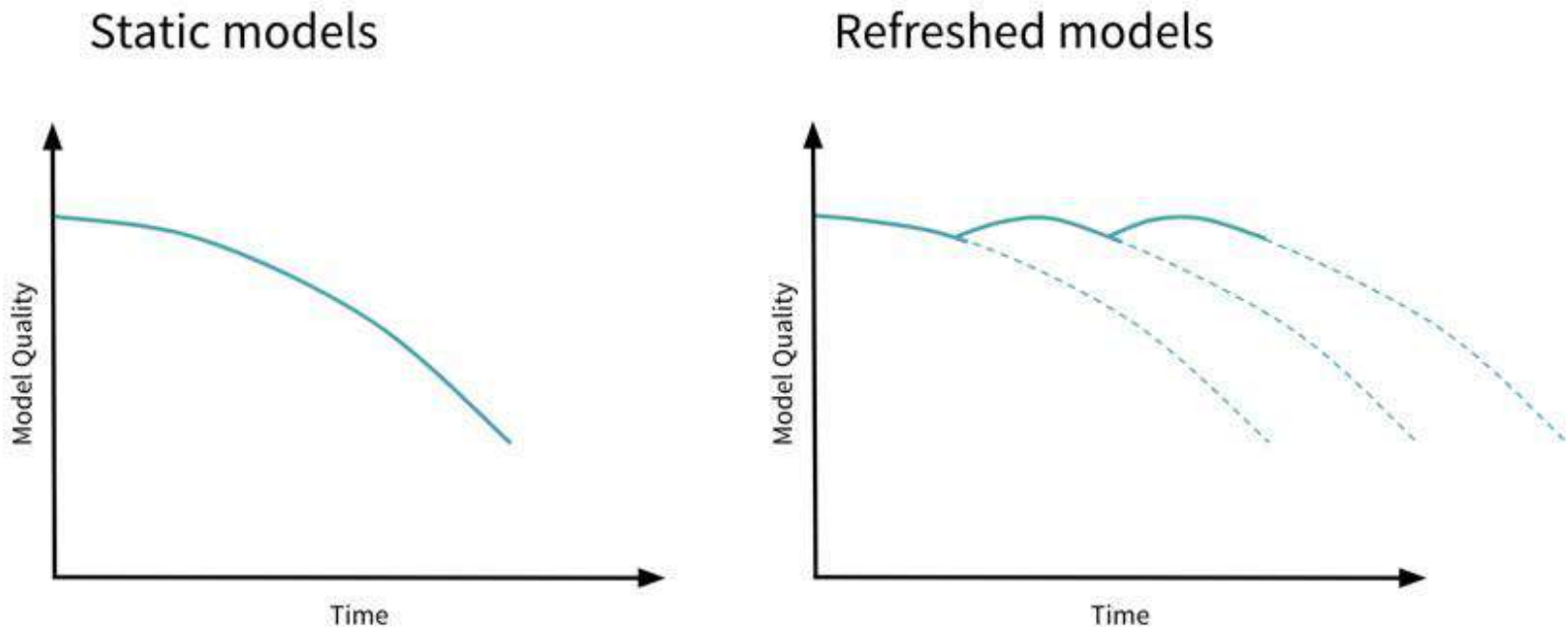
Model

Model

Model Serving

Model Serving API
Kubernetes Engine

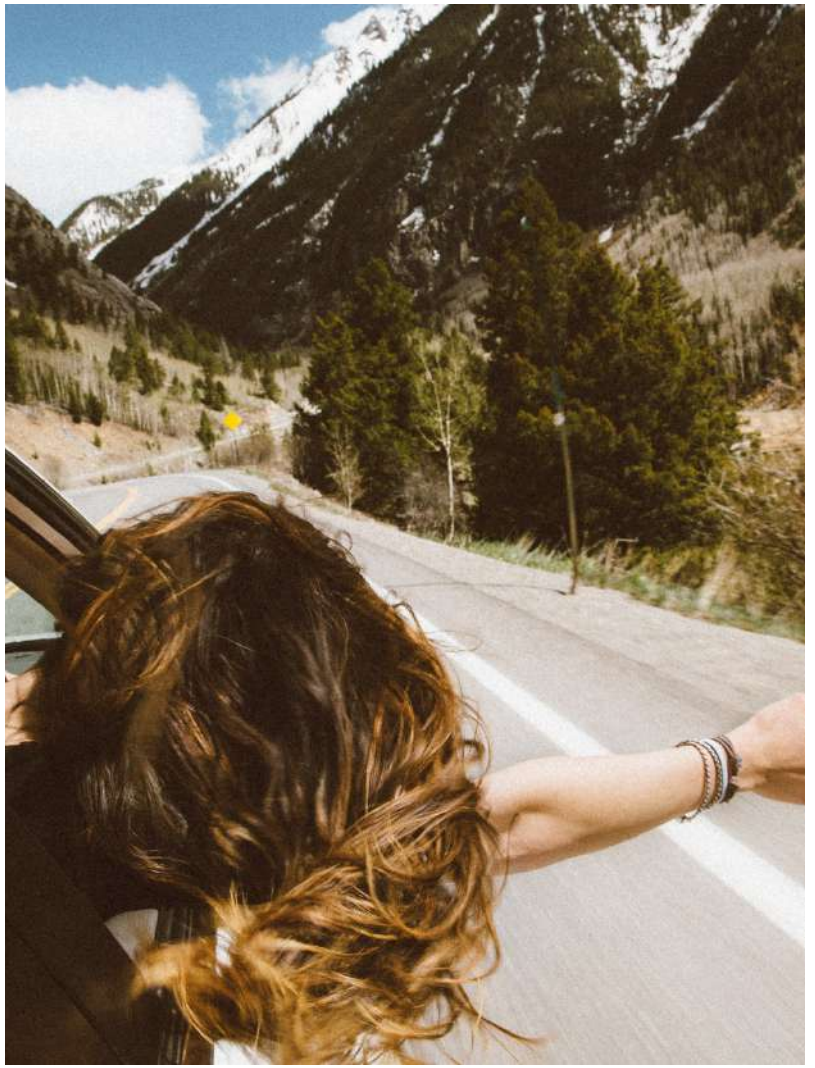
Monitoring, the last mile



Monitoring, the last mile

Sum up

- Cabify loves journeys!!!
- Data Scientists not only do theory
- How knowledge is leveraged
  - Knowledge repo as key to share
  - Lykeion more than ML platform
- Features as key transformers
- How easy is to put models in production with a the right tools



SINGULARITY
TECH DAY

#STechDay2020

THANKS AND...

SEE YOU SOON!

ORGANIZATION

plain concepts

SPONSORS

Microsoft

SUPPORT

World Central Kitchen